

生命科学实验指南系列



Metagenomics
Methods and Protocols

元基因组学：方法和步骤
（影印版）

〔德〕斯特赖特 W.R. 丹尼尔 R. 著


科学出版社

生命科学实验指南系列·典藏版



图解微生物实验指南
免疫学技术及其应用
生物衰老：研究方法与实验方案
精编细胞生物学实验指南
植物蛋白质组学实验指南
蛋白质纯化指南（原书第二版）
环境基因组学实验指南
实验动物血液生理生化参考手册
生理学实验指南
精编免疫学实验指南
酵母遗传学方法实验指南
人工细胞培养
抗体制备及使用实验指南
病毒的电子显微学研究
植物生物学与生态学实验
神经生物学实验原理与技术
DNA微阵列实验指南
基因转移：DNA和RNA的转运与表达
生物实验室管理手册（原书第二版）
精编人类遗传学实验指南
单分子技术实验指南
现代蛋白质工程实验指南
活细胞成像（原书第二版）
遗传变异分析实验指南
表皮细胞实验指南
分子克隆实验指南（原书第三版）（上下册）
精编分子生物学实验指南（原书第五版）
现代神经科学研究技术
生命科学实验设计指南
现代生物化学与分子生物学仪器与设备
分子细胞遗传学——技术和应用
精编蛋白质科学实验指南
实验细胞资源的描述标准与管理规范
实验动物设施运行管理指南
元基因组学：方法和步骤（影印版）
现代工业微生物学实验技术
真核生物转录调控——概念策略与技术（原书第二版）
动物细胞培养——基本技术和特殊应用指南（原书第六版）



科学出版中心 生物分社
联系电话：010-64012501
E-mail：lifescience@mail.sciencep.com
网址：http://www.lifescience.com.cn
销售分类建议：分子生物技术



赛拉艾芙
生命科学订阅号



定价（全套）：4500.00元

生命科学实验指南系列·典藏版

元基因组学：方法和步骤

Metagenomics: Methods and Protocols

（影印版）

〔德〕斯特赖特 W. R.　丹尼尔 R.　著

科学出版社

北京

图字：01-2012-1769 号

内 容 简 介

“生命科学实验指南系列”图书均出自名家，包括众多从 Cold Spring Harbor Laboratory Press 和 John Wiley & Sons 等国际知名出版社引进的实验室必备工具书，是生命科学领域最先进、实用、权威的实验手册类优秀图书。该系列图书简单明了，囊括了全世界最著名的生物类实验室操作方法，无论是初学者还是需要深入研究的科研工作者都能从中获益。该系列图书在读者群中有较高的知名度和美誉度，特别是以《分子克隆实验指南》和《精编分子生物学实验指南》为代表，堪称经典，分别被喻为生命科学领域的“蓝宝书”和“红宝书”。现挑选其中的精品集结成典藏版。

图书在版编目（CIP）数据

生命科学实验指南系列：典藏版/雷东锋等编著. —北京：科学出版社，2016
ISBN 978-7-03-047486-5

Ⅰ.①生… Ⅱ.①雷… Ⅲ. ①生命科学–实验–指南 Ⅳ.①Q1-0

中国版本图书馆 CIP 数据核字(2016)第 043878 号

责任编辑：王 静 李 悦
责任印制：张 伟 / 封面设计：刘新新

科学出版社 出版
北京东黄城根北街 16 号
邮政编码: 100717
http://www.sciencep.com

北京厚诚则铭印刷科技有限公司 印刷
科学出版社发行 各地新华书店经销
*
2016 年 7 月第 一 版 开本：787×1092 1/16
2016 年 7 月第一次印刷 印张：1310 1/2
字数：31 074 000

定价：4500.00 元

(如有印装质量问题，我社负责调换)

Methods in Molecular Biology™

Series Editor
John M. Walker
School of Life Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For other titles published in this series, go to
www.springer.com/series/7651

Metagenomics

Methods and Protocols

Edited by

Wolfgang R. Streit

Microbiology & Biotechnology Klein Flottbek, University of Hamburg, Hamburg, Germany

Rolf Daniel

Institute of Microbiology and Genetics, Georg August University Göttingen, Göttingen, Germany

Humana Press

Editors
Wolfgang R. Streit
Microbiology & Biotechnology
Biocenter Klein Flottbek
University of Hamburg
Hamburg
Germany
wolfgang.streit@uni-hamburg.de

Rolf Daniel
Institute of Microbiology and Genetics
Department of Genomic and Applied Microbiology
Göttingen Genomics Laboratory
Georg August University Göttingen
Göttingen
Germany
rdaniel@gwdg.de

ISSN 1064-3745 e-ISSN 1940-6029
ISBN 978-1-60761-822-5 e-ISBN 978-1-60761-823-2
DOI 10.1007/978-1-60761-823-2
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2010934277

Printed on acid-free paper

Humana Press is a part of Springer Science+Business Media (www.springer.com)

Preface

Metagenomics is a key technology to explore the DNAs from not-yet-cultivated microbes in their natural habitats. Theoretically, the microbial DNA isolated from an environmental sample represents the collective DNA of all the indigenous microorganisms and is named the metagenome. Metagenomes can be quite diverse, and, depending on the microbial community analyzed, several hundred up to several thousand different species and genomes can be present in a single metagenome. Typically, soil metagenomes are rather complex with several thousand species present, while microbial communities growing under extreme conditions (i.e., hot springs) are usually rather limited in their complexity and biodiversity. The primary goal of metagenomics is to explore this almost unlimited biodiversity. The last 10 years have already paved the way for the culture-independent assessment and exploitation of complex microbial populations for basic and applied research. Metagenomics has been defined as function-based or sequence-based cultivation-independent analysis of the collective microbial genomes present in an environment. The developed metagenomic technologies are used to complement or replace culture-based approaches and bypass some of their inherent and well-known limitations.

Besides identification of new biomolecules, metagenomics has proven to be a powerful tool for exploring the ecology, metabolic profiling, and comparison of complex microbial communities. Profiling the functions encoded by a microbial community rather than the types of organisms producing them provides a means to distinguish environmental samples on the basis of the functions selected for by the local environment and reveals insights into features of that environment. Another application of metagenomics is the genomic characterization of uncultivated microorganisms and complex communities. In addition, large-scale sequencing approaches of metagenomic DNA have been applied to reconstruct genome fragments and near-complete genomes from uncultivated species and natural consortia.

The main application area of metagenomics is the mining of metagenomes for genes encoding novel biocatalysts and drug molecules for bioindustries. Due to the complexity of most metagenomes, new sensitive and efficient high-throughput screening techniques that allow for fast and reliable identification of genes encoding suitable biocatalysts from complex metagenomes have been invented. Screens of metagenomic libraries have been based either on nucleotide sequence (sequence-driven approach) or on metabolic activity (function-driven approach).

This current book gives an overview and introduction to basic methods commonly used in laboratories that have a strong background in microbial metagenomics. All chapters are written by experts in the field, and our goal is that this book serves those who are interested in establishing metagenomics in their laboratories as a manual. Within the book, we have tried to address all working steps involved in this field: Starting from the DNA isolation from soils and marine samples to the construction and screening of the libraries, and finally we offer some advise with respect to the bioinformatic tools available to screen large sequences. An overview on strategies involved in the isolation of DNAs from environmental samples is given in the first four chapters together with the main strategies that

are currently used for the construction of metagenome libraries. Chapters 5–8 describe protocols linked to the expression of metagenome libraries in different host strains. Those include simple protocols for the construction of a library in broad host range vectors but also rather sophisticated protocols to handle *Sulfolobus* as a host strain. Furthermore, the book contains a significant number of chapters that describe a wide variety of screening technologies used for the identification of different enzymes or other biomolecules using function- and sequenced-based technologies. Altogether, the 15 chapters describe a diverse range of screening protocols for metagenome libraries. In our view, this is a very complete description of available screening protocols for all major biocatalysts and allows an easy setup of these screens in any microbiology lab.

Wolfgang R. Streit
Rolf Daniel

Contributors

ANGEL ANGELOV • *Department of Microbiology, Technical University Muenchen, Freising, Germany*
JOZEF ANNÉ • *Department of Microbiology and Immunology, Rega Institute for Medical Research, K.U. Leuven, Leuven, Belgium*
DAVID L. BALKWILL • *University of Central Florida College of Medicine, Orlando, FL, USA*
ANNALISA BALLOI • *Dipartimento di Scienze e Tecnologie Alimentari e Microbiologiche, Università degli Studi di Milano, Milan, Italy*
ANA BELOQUI • *CSIC, Institute of Catalysis, Madrid, Spain*
UWE T. BORNSCHEUER • *Department of Biotechnology and Enzyme Catalysis, Institute of Biochemistry, Greifswald University, Greifswald, Germany*
DOMINIQUE BÖTTCHER • *Department of Biotechnology and Enzyme Catalysis, Institute of Biochemistry, Greifswald University, Greifswald, Germany*
MINDY G. BROWN • *Biomedical Sciences, Florida State University College of Medicine, Tallahassee, FL, USA*
JONG-CHAN CHAE • *Division of Biotechnology, College of Environmental and Bioresource Sciences, Chonbuk National University, Iksan, Korea*
TREVOR C. CHARLES • *Department of Biology, University of Waterloo, WaterlooON, Canada*
YIN CHEN • *Department of Biological Sciences, University of Warwick, Coventry, UK*
DONALD A. COWAN • *Institute for Microbial Biotechnology and Metagenomics, University of the Western Cape, Cape Town, South Africa*
DANIELE DAFFONCHIO • *Dipartimento di Scienze e Tecnologie Alimentari e Microbiologiche, Università degli Studi di Milano, Milan, Italy*
ROLF DANIEL • *Institute of Microbiology and Genetics, Georg August University Göttingen, Göttingen, Germany*
FRANCESCA DE FERRA • *Division R&M, Eni, San Donato MilaneseMilan, Italy*
THOMAS DREPPER • *Research Centre Juelich, Institute of Molecular Enzyme Technology, Heinrich-Heine-University Duesseldorf, Juelich, Germany*
MARC G. DUMONT • *Department of Biological Sciences, University of Warwick, Coventry, UK*
MANUEL FERRER • *CSIC, Institute of Catalysis, Madrid, Spain*
MICHAEL W. FRIEDRICH • *Max Planck Institute for Terrestrial Microbiology, Marburg, Germany*
WOLFGANG GÄRTNER • *Max Planck Institute for Bioinorganic Chemistry, Mülheim, Germany*
PETER N. GOLYSHIN • *School of Biological Sciences, Bangor University, Gwynedd, UK*

JOSÉ E. GONZÁLEZ-PASTOR • *Laboratorio de Ecología Molecular, Centro de Astrobiología (CSIC-INTA), Madrid, Spain*
MARÍA-EUGENIA GUAZZARONI • *CSIC, Estación Experimental del Zaidín, Granada, Spain*
CRISTIAN GURGUI • *Kekulé-Institut für Organische Chemie und Biochemie, Rheinische Friedrich Wilhelms University Bonn, Bonn, Germany*
NELE ILMBERGER • *Microbiology & Biotechnology, Biocenter Klein Flottbek, University of Hamburg, Hamburg, Germany*
KARL-ERICH JAEGER • *Research Centre Juelich, Institute of Molecular Enzyme Technology, Heinrich-Heine-University Duesseldorf, Juelich, Germany*
ANDREW W.B. JOHNSTON • *School of Biological Sciences, University of East Anglia, Norwich, Norfolk, UK*
GEUN-JOONG KIM • *Department of Biological Sciences, College of Natural Sciences, Chonnam National University, Youngbong-Dong, Buk-Gu, Korea*
STEFAN KURTZ • *Center for Bioinformatics, University of Hamburg, Hamburg, Germany*
CHRISTIAN LEGGEWIE • *evocatal GmbH, Düsseldorf, Germany*
WOLFGANG LIEBL • *Department of Microbiology, Technical University Muenchen, Freising, Germany*
CAROLIN LÖSCHER • *Institute for General Microbiology, Christian Albrechts University Kiel, Kiel, Germany*
MASSIMO MARZORATI • *Dipartimento di Scienze e Tecnologie Alimentari e Microbiologiche, Università degli Studi di Milano, Milan, Italy*
TRACY MEIRING • *Institute for Microbial Biotechnology and Metagenomics, University of the Western Cape, Cape Town, South Africa*
REBEKKA METZGER • *Institute for General Microbiology, Christian Albrechts University Kiel, Kiel, Germany*
QUINTON MEYER • *Department of Genetics, University of Pretoria, Pretoria, South Africa*
SALVADOR MIRETE • *Laboratorio de Ecología Molecular, Centro de Astrobiología (CSIC-INTA), Madrid, Spain*
ELIZABETH H. MITCHELL • *Department of Cell Biology, University of Alabama, Birmingham, AL, USA*
KENTARO MIYAZAKI • *Institute for Biological Resources and Functions, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan*
INONGE MULAKO • *Institute for Microbial Biotechnology and Metagenomics, University of the Western Cape, Cape Town, South Africa*
J. COLIN MURRELL • *Department of Biological Sciences, University of Warwick, Coventry, UK*
JOSH D. NEUFELD • *Department of Biology, University of Waterloo, WaterlooON, Canada*
RICARDO F. NORDESTE • *Department of Biology, University of Waterloo, WaterlooON, Canada*
PHIL M. OGER • *Laboratoire de Sciences de la Terre, Ecole Normale Supérieure, Lyon, France*

SOO-JE PARK • *Department of Microbiology, College of Natural Science, Chungbuk National University, Cheongju, Korea*
SO-YOUN PARK • *Department of Biological Sciences, College of Natural Sciences, Chonnam National University, Youngbong-DongBuk-Gu, Korea*
GOPAL P. PATHAK • *Max Planck Institute for Bioinorganic Chemistry, Mülheim, Germany*
JÖRN PIEL • *Kekulé-Institut für Organische Chemie und Biochemie, Rheinische Friedrich Wilhelms University Bonn, Bonn, Germany*
SUNG-KEUN RHEE • *Department of Microbiology, College of Natural Science, Chungbuk National University, Cheongju, Korea*
MARLEN SCHMIDT • *Department of Biotechnology and Enzyme Catalysis, Institute of Biochemistry, Greifswald University, Greifswald, Germany*
RUTH SCHMITZ • *Institute for General Microbiology, Christian Albrechts University Kiel, Kiel, Germany*
DAVID J. SCHMITZ-HÜBSCH • *Department of Periodontology, University Hospital Münster, Münster, Germany*
CAROLA SIMON • *Institute of Microbiology and Genetics, Georg August University Göttingen, Göttingen, Germany*
WOLFGANG R. STREIT • *Microbiology & Biotechnology, Biocenter Klein Flottbek, University of Hamburg, Hamburg, Germany*
MARIA A. TRAINER • *Department of Biology, University of Waterloo, WaterlooON, Canada*
SONJA CHRISTINA TROESCHEL • *Research Centre Juelich, Institute of Molecular Enzyme Technology, Heinrich-Heine-University Duesseldorf, Juelich, Germany*
MARLA I. TUFFIN • *Institute for Microbial Biotechnology and Metagenomics, University of the Western Cape, Cape Town, South Africa*
TAKU UCHIYAMA • *Institute for Biological Resources and Functions, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan*
STÉPHANE UROZ • *Interactions arbres microorganismes, INRA, Champenoux, France*
LIEVE VAN MELLAERT • *Department of Microbiology and Immunology, Rega Institute for Medical Research, K.U. Leuven, Leuven, Belgium*
JOSÉ M. VIEITES • *CSIC, Institute of Catalysis, Madrid, Spain*
KRISTOF VRANCKEN • *Department of Microbiology and Immunology, Rega Institute for Medical Research, K.U. Leuven, Leuven, Belgium*
NANCY WEILAND • *Institute for General Microbiology, Christian Albrechts University Kiel, Kiel, Germany*
MARGARET WEXLER • *School of Biological Sciences, University of East Anglia, Norwich, Norfolk, UK*

Contents

Chapter 1

Molecular Methods to Study Complex Microbial Communities

José M. Vieites, María-Eugenia Guazzaroni, Ana Beloqui, Peter N. Golyshin, and Manuel Ferrer

Abstract

Microbes, which constitute a major fraction of the total biomass, are the main source of biodiversity on our Planet and play an essential role in maintaining global processes, which ultimately regulate the functioning of the Biosphere. Recent emergence of "metagenomics" allows for the analysis of microbial communities without tedious cultivation efforts. Metagenomics approach is analogous to the genomics with the difference that it does not deal with the single genome from a clone or microbe cultured or characterized in laboratory, but rather with that from the entire microbial community present in an environmental sample; it is the community genome. Global understanding by metagenomics depends essentially on the possibility of isolating the entire bulk DNA and identifying the genomes, genes, and proteins more relevant to each of the environmental sample under investigation. Following on this, in this chapter, we provide an analysis of methods available to isolate environmental DNA and to establish metagenomic libraries that can further be used for extensive activity screens.

Key words: Metagenomics, Cosmid, Fosmid, Phage library, Screening

1. Introduction

Microbes, the most abundant organisms on Earth, play a major role in maintaining global element cycling processes and facilitating the self-sustainable functioning of the Biosphere. From this point of view, it is crucial to generate a thorough understanding of these key microorganisms and processes they facilitate. However, at present, we simply do not know the extent of the functional diversity that microbes encompass: a classical theoretical analysis endeavors a population of prokaryotes on Earth of about 10^{30} bacteria, few order of magnitude higher than the

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668, DOI 10.1007/978-1-60761-823-2_1,

number of stars in the known Universe (estimated 10^{22}–10^{24}) (1–3), with most microbes being members of complex communities. Invertebrate guts are certainly one of the most dense and diverse niches [10^{9}–10^{11} cells per mL of gut fluid (4)], followed by soil [10^{7}–10^{9} cells per gram (5)], and oligotrophic superficial sea- and freshwater [10^{5}–10^{6} bacteria per milliliter (6)]. Any individual survey to study such diversity is limited due to the relatively poor capacity of growth of most microorganisms that is offered even by rather sophisticated resources available for culturing (7). To circumvent this problem, a wide range of approaches collectively described as "metagenomics" have been developed to study communities through the analysis of their genetic material without culturing individual organisms (8). Metagenomics is analogous to genomics with the difference that it does not deal with the single genome from a clone or microbe cultured or characterized in laboratory, but rather with that from the entire microbial community present in an environmental sample; it is the community genome. Metagenomics represents a strategic concept that includes investigations at three major interconnected levels (sample processing, DNA sequencing, and functional analysis), with an ultimate goal of getting a holistic view of the functioning of microbial World. While many of the technical limitations to processing of samples have been overcome in the last decade (multiwell DNA extractions, single-cell isolation, sequence analysis by technologies such as 454 or Solexa platforms), we believe that the major hurdles still are (1) the adequate metagenome coverage, since genes of different organisms are be present in very different concentrations in the DNA used to construct the libraries or for sequencing, (2) the integrating and filtering gene sequences and experimental evidences to facilitate functional assignments of unknown genes, organisms, and communities and to recreate functional networks, and (3) the computational aspects of data archiving, analysis ,and visualization of vast numbers of DNA sequences which are released to databases. In this respect, lessons from 20 years of metagenomics and four of high-throughput DNA sequencing [first analyses of microbial communities through massive sequencing were published in 2004 (9, 10)] tell that giga-base amounts of environmental sequences can easily be generated to a large extent, but only a fraction of them can properly be annotated in terms of gene functions (~50% of the potential protein-coding genes lacked any functional assignment). More importantly, DNA sequences per se are not that helpful in linking genes to specific functions as we know that more than 60% of genes are ubiquitous and have similar housekeeping functions in different organisms. Therefore, in this chapter, we try to provide a broad view on current technical issues to illustrate the potential of getting appropriate metagenomic material to create representative gene libraries, as the first step for analysing community genomes.

Following the above observations, the principal measure of phylogenetic relatedness, and thus of biodiversity, is the sequence of the 16S ribosomal RNA gene in prokaryotes and its equivalent 18S rRNA gene in eukaryotes. Determination of very large numbers of such sequences has revealed that natural environments contain vast numbers of diverse microorganisms, but only a fraction of them can properly be analyzed (11). This "great plate count anomaly" (12), in fact, observed from early 1930s stimulated the development of new efficient tools to circumvent problems linked to the cultivation of microbes in artificial medium, the so-called metagenomics (13). These are often described as culture-independent approaches and, in terms of the organisms being accessed and mined, this is the case. However, the need for large amounts of cell biomass for gene and genomic analysis always requires cultivation of a producer microbe, except for DNA sequencing which requires direct separation of cells and bulk DNA. The difference here is that cultivation refers to that of a surrogate organism, the host exploited as a reservoir for archiving the harvested genetic resources. Considering these requirements, metagenomics is often based on a general strategy of producing large amount of environmental DNA to achieve two goals: (1) discovery of new gene sequences coding for enzymes and drugs and (2) random sampling and archiving of the genomes from a small subset of organisms present in an environment for subsequent in silico analysis (14, 15). Both research windows are essential as the microorganisms are known to be the "gatekeepers" of environmental processes. However, it is essential to note that the relative abundance of representatives of a certain group of microbes is not necessarily linked to the importance of that group in the community functioning: common organisms may not necessarily play a critical role in a community despite their numbers, and organisms that only muster 0.1% fraction (e.g., nitrogen fixers) can be of pivotal importance. What this means, in terms of microbial ecology, is that the structural and functional information based on more reductionist approach, that is, classical functional genomics based on single organisms, may not provide appropriate understanding of complex communities.

To date, much of the research has been focused on bulk DNA sequencing. The analysis of samples at sequence level somehow has lower resolution but can access much greater genomic information of untapped microbial biodiversity in many environments. In contrast, the second approach shows better options to link specific microbes to specific ecological functions. In one of the first examples, the Sargasso Sea genome sequencing project, the authors performed a size-selective filtration for enrichment of the microbes of a certain size (10). Actually, new developments involve the direct separation of cells or preferably the enrichment using ^{13}C-labeled compound directly related to primary ecological functions (16). A particular elegant strategy combines the extraction of almost

complete genomes from uncultured microbes in complex communities (with up to 5,000 species) by high-resolution stable isotope probing (SIP) to reconstruct their metabolisms and to link specific microbes, whose DNA is separated by ultracentrifugation, to specific ecological functions (17). Here, authors provided a good genome coverage of dominant organisms under dynamic utilization of different nutrients and were able to link environment-specific organisms and processes that are catalyzed by these microbes. However, despite their great potential, the main drawback of enrichment methodologies is the danger of a nonproportional accumulation of fast-growing microbes in the community, which is not necessarily relevant to the native ecosystem, followed by the reduction of the natural diversity in the sample.

We should also point out that metagenomics is not limited to prokaryotes: eukaryotic microbial diversity is also enormously diverse and are hence of a great interest for exploration of their functional diversity. Owing to the problem of introns in eukaryotes, considerable effort has been invested in the isolation of RNA and its conversion to cDNA, rather than dealing with genomic DNA. This requires isolation of full-length mRNAs, reverse transcribing them, and analysis of the cDNA libraries. Here, the RNA extraction technique is critical, since it needs to extract RNA from thick-walled cells of fungi and yeasts, and their spores. Further, as a complement to the long-standing trend towards reductionism, metagenomics seeks to treat the community as a whole. However, this is not an easy task, specially for sample processing, as we know that environmental samples also contain picoeukaryotes (size $<2–3\ \mu m$) whose population composition varies dynamically in response to both seasonal and spatial gradients in environment (3, 18). Therefore, a general strategy for sample processing is recommended for metagenomics analysis in the future, in which parts of microbial communities are processed separately, using single microdroplets and cell-free translation systems together with cell sorting ("single-cell genomics"), accompanied by the integration these data with those obtained using mixed microbial communities (19–21). Finally, we should consider that genome coverage is an ephemeral term in complex communities, since individual community members are be present in varying numbers in a sample and their genomes are extracted with different efficiencies (see Note 1). Therefore, the genes of different organisms will be present at very different concentrations in the DNA used as material to construct the libraries or to sequencing. For this reason, attempts to obtain (or even calculate the size of) a sample providing good coverage of all genomes present in a sample are rare and limited to samples from extreme environments (22), known to contain microbial communities of very limited complexity/diversity. Further advances in this field are demanded to appropriately reconstruct the metabolism in more complex

microbial communities. Below, we list a series of protocols for extraction of environmental DNA and further production of metagenomic libraries.

2. Materials

2.1. Sampling

1. 500-kDa NMWL ultrafiltration disc (Biomax polyethersulfone, Millipore).
2. Filtration device Pellicon TFF 0.1 μm (Millipore™).
3. 500-kDa filtration disc Amicon® system (Millipore).
4. TE buffer pH 8.0.
5. Disruption buffer: 0.2 M NaCl, 50 mM Tris–HCl, pH 8.0.
6. Nycodenz (0.8–1.3 g/mL in distilled H_2O).
7. Phosphate-buffered saline (PBS) buffer: 137 mM NaCl, 2.5 mM KCl, 10 mM phosphate, pH 7.4.

2.2. DNA Extraction

1. Lysozyme solution (10% w/v).
2. RNase solution (1% w/v) free of DNase.
3. Proteinase K (1% w/v).
4. SDS solution (10% w/v).
5. Cetyltrimethylammonium bromide (CTAB) solution.
6. 10% w/v in 0.7 M NaCl.
7. $ChCl_3$:Isoamyl alcohol.
8. Isopropanol.
9. Ethanol (70% v/v).
10. TE buffer, pH 8.0.
11. DNA Clean & Concentrator from Zymo Research Corp.

2.3. DNA Isolation with Kits

1. UltraClean MegaPrep (MoBio Laboratories, Inc.).
2. G'NOME® DNA Extraction Kit (BIO101).

2.4. Nucleic Acid Quantification

1. Quant-iT™ PicoGreen® Kit.

2.5. 16S rRNA Gene Libraries Construction

1. pGEMT-Easy (Promega).
2. T4 DNA ligase buffer (10×).
3. T4 DNA ligase.
4. Primer 16F530 (5′-TTCGTGCCAGCAGCCGCGG-3′).
5. Primer 16R1492 (5′-TACGGYTACCTTGTTACGACTT-3′).
6. *Escherichia coli* DH5α.

2.6. Metagenomic Libraries Construction

2.6.1. In pCCFOS Vector 1

1. CopyControl™ Fosmid Library Production kit (EPICENTRE).
2. GELase (EPICENTRE) kit.
3. Microcon-100 (Millipore) concentrator membrane (100 kDa cut-off).
4. 3 M Sodium Acetate (pH 7.0).
5. TE buffer, pH 8.0.
6. 70% ethanol and 100% ethanol.
7. PD buffer: 10 mM Tris–HCl pH 8.3, 100 mM NaCl, 10 mM $MgCl_2$.
8. *E. coli* EPI300-T1^R.
9. Chloramphenicol 12.5 mg/mL.
10. LB (Luria–Bertani) medium.

2.6.2. In pLAFR3 Cosmid

1. Large-Construct Kit (Qiagen).
2. Tetracycline (Tc) 10 μg/mL.
3. ATP-dependent exonuclease.
4. *Hin*dIII, *Eco*RI, *Sau*3AI, and *Bam*HI enzymes.
5. Shrimp Alkaline Phosphatase (SAP; from Biotec ASA).
6. Microcon-100 (Millipore).
7. GeneClean Kit (BIO101).
8. Gigapack XL (Stratagene).
9. NEB3 and 1 (New England Biolabs Buffers 1 and 3).
10. 0.5 M ethylenediaminetetraacetic acid (EDTA) pH 8.
11. GELase (EPICENTRE).
12. Bovine Serum Albumin (BSA) – nuclease free (10×).
13. Ligation Buffer NEB1 (10×).
14. T4 DNA Ligase (New England Biolabs).
15. SM buffer: 50 mM Tris–HCl pH 7.5, 0.1 M NaCl, 8.5 mM $MgSO_4$, and 0.01% (w/v) gelatin.
16. Chloroform.
17. *E. coli* DH5α or XL1Blue.
18. 1 M $MgSO_4$.
19. 2% (w/v) maltose.
20. LB medium.

2.6.3. In Lambda Zap® Express System

1. Zap Express pBK vector (Stratagene).
2. *Sau*3AI enzyme.
3. 10× BSA.
4. NEB1 (New England Biolabs Buffer 1).

5. 0.5 M EDTA pH 8.
6. *E. coli* XL1 MRF′.
7. Gigapack XL (Stratagene).
8. GELase (EPICENTRE).
9. 1 M $MgSO_4$.
10. 2% (w/v) maltose.
11. NZY soft agar.
12. Chloroform.
13. LB medium.

2.6.4. Bacterial Chromosomal DNA in InCert® Agarose Gel Plugs

1. Polyethylenglycol.
2. Deoxycholate.
3. RNase A (20 μg/mL).
4. N-Lauroylsarcosine.
5. Proteinase K.
6. Lysozyme (1 mg/mL).
7. 10× BSA.
8. EC lysis buffer: Combine 6 mM Tris–HCl pH 7.6, 1 M NaCl, 0.1 M EDTA pH 8.0, 0.5% polyethylenglycol, 0.2% deoxycholate, and 0.5% N-lauroylsarcosine.
9. PETT IV buffer: 10 mM Tris–HCl pH 7.6, 1 M NaCl.
10. ESP buffer: 0.5 mM EDTA pH 8.0, N-Lauroylsarcosine (1%), Proteinase K (1 mg/mL).
11. ES buffer: 0.5 mM EDTA, pH 8.0, N-Lauroylsarcosine (1%).
12. TE buffer: 10 mM Tris, pH 7.5, 1 mM EDTA.
13. TE buffer plus phenylmethylsulfonyl fluoride (PMSF): 10 mM Tris pH 7.5, 0.1 mM EDTA, 1 mM PMSF.
14. PFG-TBE buffer: 100 mM Tris pH 8.0, 100 mM boric acid, 0.2 mM EDTA.
15. 2× LB broth.

2.6.5. BAC Library Construction

1. 10× Ligation buffer.
2. pBeloBAC11 vector.
3. 10 mM ATP.
4. T4 DNA ligase.
5. *Bam*HI enzyme.
6. NEB1 (New England Biolabs Buffer 1).
7. 1-Butanol.
8. Epicentre IndigoBac kits.

2.7. Activity Screens

2.7.1. Esterase Screens

1. 50 mM HEPES Buffer pH 7.5, 0.4% (w/v) agarose.
2. Fast Blue RR solution in dimethyl sulfoxide (DMSO; 80 mg/mL).
3. α-Naphtyl acetate solution in acetone (20 mg/mL).

2.7.2. Cellulase-Like Screens

1. 0.5% Congo red.

2.7.3. P450 Oxidoreductase Screens

1. 100 mL 50 mM Tris–HCl pH 7.5, 0.4% agarose, *p*-nitrophenoxycarboxylic acids such as 12-*p*NC10 in DMSO (15 mM).

2.7.4. Laccase-Like Screens

1. 0.4% Soft agar, 50 μM syringaldazine (SGZ).
2. 0.1 M Sodium phosphate buffer pH 8, 50 μL 0.05 M 2, 6-dichloro-phenolindophenol, 100 μL 0.15 M NAD solution.
3. 0.005 M 5-methyl-phenazinium methyl sulfate.

2.7.5. Screens for Sugar Fermenting Enzymes Able to Produce Alcohol

1. 0.4% agarose made up in a 0.1 M sodium phosphate buffer pH 8.
2. 0.05 M 2,6-dichlorophenolindophenol.
3. 0.15 M NAD solution.
4. 0.005 M 5-methyl-phenazinium methyl sulfate.

2.7.6. Alcohol Oxidoreductase Screens

1. 1,2-ethanediol, 2,3-butanediol, or a mixture of 1,2-propanediol and glycerol.
2. Pararosaniline: 2.5 mg/mL of 95% ethanol; not autoclaved.
3. Sodium bisulfite (unsterilized dry powder).

2.7.7. Catechol Dioxygenase

1. 50 mM Tris–HCl pH 7.5.
2. 10 mM Catechol in methanol.

2.7.8. Polyol Oxidase-Like

1. Methyl-β-D-galactopyranoside.
2. 50 mM Sodium phosphate solution, pH 7.0.
3. Catalase (Sigma).
4. 0.5 mM $CuSO_4$.

3. Methods

Several studies showed that the size of DNA fragments extracted from environmental samples varied in a range between less than 10 kb and more than 400 kb (see Note 2), depending on the sample and the mechanical, chemical, or enzymatic protocols used for the DNA extraction (23). Some samples need an enrichment approach combined with a cell separation and gradient

centrifugation to isolate high-molecular-weight DNA. Also, in many cases, the extraction of inhibitor-free metagenomic DNA (i.e., from polluted sediments) is required and, in this context, humic compounds removal/absorption and ion-exchange treatments have been recommended (24). The DNA extraction procedure (together with its further size sorting) is a critical step and considerably differs when constructing large-insert libraries (i.e., fosmid, cosmid, or pBAC libraries constructed for archiving and gene probing screens) and small-insert expression libraries (e.g., those in lambda phage and plasmid vectors, specially aimed at the activity screening). First of all, one could divide samples depending on the source, that is, liquid (e.g., marine environments) or solid samples (e.g., soil, marine or river sediments). In subsequent examples, we describe key strategies to isolate high-quality DNA to construct metagenomic libraries (see Note 3).

3.1. Samples Handling

3.1.1. Superficial Sea- and Freshwater Samples Treatment

1. Immediately after collection, samples (approximately 2–20 L) are filtered onto a 500-kDa NMWL ultrafiltration.
2. After this, the filters are cut into strips (1 cm × 2 mm) and used directly for DNA extraction (see the protocols described below). Alternatively, in the case of large sample amounts (e.g., 100–200 L), a tangential flow filtration device can be used, such as Pellicon TFF 0.1 μm to separate solid particles and picoeukaryotic organisms (2–4 μm).
3. Part of the overconcentrated product (retentate solution, e.g., 10 mL) can be used for enrichment of cells with a desired supplement (for example, minimal medium supplemented with petroleum, to enrich cells able to metabolize it) or this retentate solution can be filtered onto a 500-kDa pore size with the Amicon® system, and the filter cut into strips and used directly for DNA extraction.

3.1.2. Solid Samples Treatment

The cloning efficiency of metagenomic fragments greatly depends on the methodology used to purify fragments of DNA. Many existing methods to isolate such fragments directly from environmental samples, especially soils and sediments, are hampered by the problems of mechanical shearing, due to physical forces (e.g., bead beating) or DNA quality (e.g., copurification with humic acids). Some samples can be used directly for DNA extraction or need more than a centrifugation step for isolation of microorganisms. In this case, a "Nycodenz" extraction technique is suggested.

3.1.2.1. Direct DNA Extraction

1. Soil or sediment samples (1–5 g) are resuspended in 5–40 mL (depending of the sample properties) of TE buffer pH 8.0 and mixed by inverting the tube 10–15 times.
2. Then samples are mixed with moderate shaking to release cells from the solid matrix, and centrifuged at low speed (approximately 200–400 × *g* for 1–5 min) to eliminate bulky soil particles.

3. Then place the supernatant in a new tube and centrifuge it at 6,000 × *g* for 15–30 min at 4°C.
4. After that, discard supernatant and conserve the pellet for subsequent DNA extraction.

3.1.2.2. Nycodenz Gradient Prior DNA Extractions

To avoid DNA damage during purification from environmental samples, a Nycodenz extraction technique is highly recommended. During the physical separation of the bacterial fraction using a Nycodenz cushion a whitish band of microbial biomass is obtained at the interface between the Nycodenz and the aqueous layer by using a high-speed density gradient. This method has been used successfully to isolate DNA from freshwater, compost, rhizosphere-associated soils, and pristine and contaminated sediments. The procedure is outlined below:

1. Prepare sample suspension: to 15 g sample add disruption buffer (35 mL total volume: 0.2 M NaCl, 50 mM Tris–HCl pH 8.0) and mix (preferably overnight with orbital shaking).
2. Centrifuge at low speed (approximately 200–400 × *g* for 1–5 min) to eliminate large soil particles and then use supernatant for biomass separation via Nycodenz.
3. 25-mL of the soil homogenate is transferred to an ultracentrifuge tube, and 9–11 mL of Nycodenz (0.8–1.3 g/mL) is carefully pipetted to form a layer below the homogenate.
4. Centrifuge at 10,000 × *g* for 20–40 min at 4°C
5. A faint whitish band containing bacterial cells is resolved at the interface between the Nycodenz and the aqueous layer. This band is transferred into a sterile tube. Note that sometimes, soils contain a lot of small particles which are not separable: they cover Nycodenz surface, making solid layer mixed with microbial biomass (this problem is typical for clay soils).
6. Approximately 35 mL of PBS buffer is added and the cells pelleted by centrifugation at 10,000 × *g* for 20 min. The cell pellet, resuspended in 0.5–2.0 mL TE buffer pH 8.0, is then ready for lysis and DNA extraction.

3.2. DNA Extraction Protocols

The following procedure is recommended when high quantity of humic acids is presented in the soil samples and for Nycodenz separated biomass to minimize the volume of solvents required for large-scale soil analyses.

3.2.1. Isolation of High-Quality DNA by Phenol:Chloroform Method Followed by DNA Cleaning

1. To the solution obtained as above (cell pellet in TE buffer), add 25 μL of Lysozyme solution (10% w/v, prepared prior to use) and incubate for 2 h at 37°C (1,500 × *g*).
2. Add 6 μL RNase solution (1% w/v) free of DNase and incubate for approximately 30 min at 37°C.

3. Add 8-μL solution Proteinase K (1% w/v) and 60-μL SDS solution (10% w/v) and incubate for 30 min at 50°C till solution becomes clear and viscous.
4. Add 100-μL NaCl 5 M and mix gently by inverting the tube 4–6 times.
5. Add 80 μL of a CTAB solution prewarmer at 65°C (10% w/v in 0.7 M NaCl), mix gently by inverting the tube 4–6 times, and incubate 10 min at 65°C.
6. The final volume should be around 748 μL.
7. Add 750-μL $ChCl_3$:Isoamyl alcohol, mix gently by inverting the tube 4–6 times, then centrifuge 3 min at 14,000 × *g*, and quickly transfer the above supernatant to a new 2-mL Eppendorf.
8. Add 350-μL $ChCl_3$:Isoamyl alcohol and 350-μL Phenol, mix gently by inverting the tube 4–6 times, then centrifuge 3 min at 14,000 × *g*, and quickly transfer the above supernatant to a new 2-mL Eppendorf.
9. Add 650-μL $ChCl_3$:Isoamyl alcohol, mix gently by inverting the tube 4–6 times, then centrifuge 3 min at 14,000 × *g*, and quickly transfer the above supernatant to a new 1.5-mL Eppendorf.
10. Finally add 0.6 volume of isopropanol, mix gently by inverting the tube 4–6 times, transfer suspension to ice during 10 min, centrifuge for 30 min at 14,000 × *g*, and eliminate the supernatant.
11. Add 500-μL ethanol (70% v/v) to the DNA pellet, mix gently by inverting the tube 4–6 times, centrifuge for 30 min at 14,000 × *g*, and eliminate the supernatant.
12. Quickly transfer the tube to speed-vac and dry for about 5 min without heating.
13. Resuspend the DNA pellet in 500-μL TE buffer, pH 8.0.
14. If quality of DNA of the previous step is good enough, then one could proceed directly to the digestion and cloning step. If not, an ultrapure DNA protocol should be applied: DNA Clean & Concentrator from Zymo Research Corp., has been shown to be effective for purification of quality DNA. By using this product, one could purify DNA by employing a single-buffer system that allows the efficient and selective DNA adsorption onto a matrix. Here, it is important to use at least 200-μg DNA, since the recovery of DNA is, in the majority of samples, lower than 40%.

3.2.2. Isolation of High-Quality DNA by Commercial Kits

Commercial kits such as UltraClean MegaPrep and G'NOME® DNA Extraction Kit may be used for isolation of metagenomic DNA from eukaryotic or prokaryotic cells and tissues in less than

2 h with no organic extractions. Preliminary separation of cellular biomass from soil homogenate via Nycodenz gradient is recommended in order to achieve maximal DNA recovery per gram of soil. DNA purification kits from other manufacturers presumably work equally well, but they have not been tested in our laboratory. The GNOME®DNA kit uses RNase Mix to eliminate RNA immediately after lysis, and Protease Mix to rapidly degrade cellular proteins. This is followed by a proprietary "salting out" technique that precludes the need for phenol, chloroform, or other organic extractions. Preparation of metagenomic DNA using this kit is described below:

1. Immediately after collection samples are either stored in 95% EtOH at 4°C or are shock-frozen in liquid nitrogen, followed by storage at −80°C.
2. Alternatively, 10-g soil is directly homogenized with 40 mL of a 0.2 M NaCl, 50 mM Tris–HCl pH 8.0 buffer, mix (overnight, 4°C, orbital shaking) and then centrifuge at low speed (approximately 200–500 × *g* for 1–5 min) to eliminate large soil particles. Cells (plus small particles) are separated by centrifugation at high speed (9,000 × *g*, 15 min). Freshwater samples (>600 mL) are subjected to cell separation either by centrifugation (9,000 × *g*, 15 min) or filtration through a 0.20-μm filter (the filter is removed and used directly). DNA was isolated from these samples following steps 2–10. To each 0.1 g of separated cells ground in liquid nitrogen, add 1.85 mL *Cell Suspension Solution* (use a 15-mL clear plastic tube for efficient mixing). Mix until the solution appears homogeneous.
3. Add 50 μL of *RNase Mix*, mix thoroughly. Add 100 mL of *Cell Lysis/Denaturing Solution* and mix well.
4. Incubate at 55°C for 15 min.
5. Add 25-μL *Protease Mix* and mix thoroughly.
6. Incubate at 55°C for 30–120 min (the longer time will result in minimal protein carry over and will also allow for substantial reduction in residual protease activity).
7. Add 500-μL *"Salt-Out" Mixture* and mix gently yet thoroughly. Divide sample into 1.5-mL tubes. Refrigerate at 4°C for 10 min.
8. Spin for 10 min at maximum speed in a microcentrifuge (at least 10,000 × *g*). Carefully collect the supernatant, avoid the pellet. If a precipitate remains in the supernatant, spin again until it is clear. Pool the supernatants in a 15-mL (or larger) clear plastic tube.
9. To this supernatant, add 2-mL TE buffer and mix. Then, add 8 mL of 100% ethanol. If spooling the DNA, add the ethanol

slowly and spool the DNA at the interphase with a clean glass rod. If centrifuging the DNA, add the ethanol and gently mix the solution by inverting the tube.

10. Spin for 15 min at 1,000–1,500 × *g*. Eliminate the excess ethanol by blotting or air-drying the DNA.
11. Dissolve the genomic DNA in TE (10 mM Tris pH 7.5, 1 mM EDTA).

3.3. Nucleic Acids Quantification with Quant-iT™ PicoGreen®

Detecting and quantitating amounts of DNA is extremely important to have adequate proportions of DNA insert and vector in a ligation reaction to generate a satisfactory metagenomic library. Quant-iT™ PicoGreen® dsDNA reagent is an ultrasensitive fluorescent nucleic acid stain for quantitating double-stranded DNA (dsDNA) in solution. It is very important to measure DNA (genomic and vector) previous to enzymatic restriction and ligation reactions that are described in the following sections. Next, we detail the protocol using this fluorescent nucleic acid stain. For that, a standard curve is prepared as follows:

1. Prepare a 2 μg/mL stock solution of dsDNA in TE buffer.
2. Determine the DNA concentration on the basis of absorbance at 260 nm (A_{260}) in a cuvette with a 1-cm pathlength; an A_{260} of 0.04 corresponds to 2 μg/mL dsDNA solution. For a standard curve, we commonly use bacteriophage lambda or calf thymus DNA, although, any purified dsDNA preparation may be used. The lambda DNA standard, provided at 100 μg/mL in the Quant-iT™ PicoGreen® Kits, can simply be diluted 50-fold in TE to make the 2 μg/mL working solution.
3. For the standard curve, dilute the 2 μg/mL DNA stock solution into disposable 96-well microtiter plates. Then, add 0.2 mL of the aqueous working solution of Quant-iT™ PicoGreen® reagent to each cuvette. Mix well and incubate for 2–5 min at room temperature, protected from light.
4. After incubation, measure the sample fluorescence using a spectrofluorometer or fluorescence microplate reader and standard fluorescein wavelengths (excitation ~480 nm, emission ~520 nm). To ensure that the sample readings remain in the detection range of the fluorometer, the instrument's gain should be set so that the sample containing the highest DNA concentration yields fluorescence intensities near the fluorometer's maximum. To minimize photobleaching effects, keep the time for fluorescence measurement constant for all samples.
5. Subtract the fluorescence value of the reagent blank from that of each of the samples. Use corrected data to generate a standard curve of fluorescence *versus* DNA concentration. For sample analysis, continue with steps 5–9.

6. Dilute the experimental DNA solution in TE to a final volume of 0.2 mL in 96-well microtiter plates. You may alter the amount of sample diluted, provided that the final volume remains 0.2 mL. A higher dilution of the experimental sample may diminish the interfering effect of certain contaminants.
7. Add 0.2 mL of the aqueous working solution of the Quant-iT™ PicoGreen® reagent to each sample. Incubate for 2–5 min at room temperature, protected from light.
8. Measure the fluorescence of the sample using instrument parameters that correspond to those used when generating your standard curve. To minimize photobleaching effects, keep the time for fluorescence measurement constant for all samples.
9. Subtract the fluorescence value of the reagent blank from that of each of the samples. Determine the DNA concentration of the sample from the standard curve generated in *DNA Standard Curve*.
10. The assay may be repeated using a different dilution of the sample to confirm the quantitation results.

3.4. 16S rRNA Gene Libraries Construction

Application of molecular ecological methods, especially those based on surveys of genes after PCR amplification, has allowed cultivation-independent investigations of the microbial communities of environmental samples to be made. In particular, it has proven to be useful and powerful markers for the presence of bacteria in samples (25, 26). The utility of these markers is facilitated by the availability of primers that allow amplification of almost complete gene or its product and by the phylogenetic inferences that can be made from the resultant nucleotide sequences, permitting a quite accurate picture of the bacterial community. Therefore, before the construction of the correspondent metagenomic library, it is important to analyze a library of PCR-amplified 16S rRNA genes to know the microorganism diversity of the sample. The procedure is outlined below:

1. Bacterial 16 rRNA genes are amplified from the environmental sample of genomic DNA using two general bacterial 16S rRNA primers 16F530 (5′-TTCGTGCCAGCAGCCGCGG-3′) and 16R1492 (5′-TACGGYTACCTTGTTACGACTT-3′) and the polymerase chain reaction (27).
2. The PCR reaction (50 μL) is performed with an annealing temperature of 50°C, and 25 cycles should be used.
3. The PCR products are purified from a 1% agarose gel and inserted into the pGEMT-Easy vector as follows:

   Reaction 1:

   1-μL pGEMT-Easy

   1-μL T4 DNA ligase buffer (10×)

0.5-μL T4 DNA ligase

3.3-μL PCR product

4.1-μL MilliQ water

Reaction 2:

1-μL pGEMT-Easy

1-μL T4 DNA ligase buffer (10×)

0.5-μL T4 DNA ligase

7.0-μL PCR product

0.5-μL MilliQ water

4. Ligate at 4°C overnight.
5. Then, the product of this ligation (2 μL) is used to transform 50-μL competent *E. coli* DH5α cells.
6. Cells are plated in LB agar Amp_{50}-XGal plates and around 100 positives random selected clones (white colonies) are sequenced using the M13f primer.

3.5. Metagenomic Libraries Construction

There are two distinct strategies taken in metagenomics, according to the primary goal. First, large-insert libraries [cosmid, fosmid, or bacterial artificial chromosomes (pBACs)] are constructed for archiving and sequence homology screening purposes: to capture the largest amount of the available genetic resources available in the sample and archive it for further studies/interrogation. Second, small-insert expression libraries, especially those made in lambda phage vectors, are constructed for activity screening. The small size of the cloned fragments means that most genes present in the appropriate orientation will be under the influence of the extremely strong vector expression signals, and thus have a good chance of being expressed and detected by activity screens. Though these two strategies may differ in some technical aspects, both are increasingly used together, due to their complementarities: activity mining often reveals novel enzymes, but the nature of the organism from which they originate can rarely be determined, nor can their genetic context, which may harbor equally or even more interesting similar or related enzymes. Primary enzyme discovery in an expression library, followed by identification of the same gene in a large-insert library and genome walking on the identified fragment, constitutes a powerful means of maximizing the discovery process and identifying the interesting new organisms that are producing such enzymes. Below, we describe the methods used to construct small and large libraries.

3.5.1. Using the pCCFOS Vector

The *CopyControl™ Fosmid Library Production* kit utilizes a strategy of cloning randomly sheared, end-repaired DNA with an average insert size of 40 kbp. Shearing the DNA into approximately 40-kb fragments leads to the highly random generation of DNA fragments

in contrast to more biased libraries that result from partial restriction endonuclease digestion of the DNA. Frequently, genomic DNA is sufficiently sheared as a result of the purification process, so additional shearing is not necessary. Test the extent of shearing of the DNA by first running a small amount of it (around 100 ng). Run the sample on a 20-cm long gel 1% agarose at 30–35 V overnight at 4°C and stain with ethidium bromide.

3.5.1.1. Partial Shearing of DNA Insert for pCCFOS Cloning

1. If 10% or more of the genomic DNA migrates with the Fosmid control DNA provided with the kit (36-kb size), then you can proceed to the end-repair protocol.
2. If the genomic DNA migrates slower (higher MW) than the 36-kb fragment, then the DNA needs to be sheared.
3. Shear the DNA (2.5 μg) by passing it through a 200-μL small-bore pipette tip. Aspirate and expel the DNA from the pipette tip 50–100 times.
4. If the genomic DNA migrates faster than the 36-kb fragment (lower MW), then it has to be sheared too much and should be reisolated.
5. For the end-repair protocol, take into account these suggestions:
6. Thaw and thoroughly mix all of the reagents listed below before dispensing; place on ice. Combine the following on ice:

   8-μL 10× End-Repair Buffer

   8-μL 2.5 mM dNTP Mix

   8-μL 10 mM ATP

   32-μL sheared insert DNA (approximately 4.3 μg) (the end-repair reaction can be scaled up or scaled down as dictated by the amount of DNA available).

   20-μL sterile water

   4-μL End-Repair Enzyme Mix

   80-μL Total reaction volume
7. Incubate at room temperature for 45 min.
8. Add gel loading buffer and incubate at 70°C for 10 min to inactivate the End-Repair Enzyme Mix.
9. Select the size of the end-repaired DNA by low melting point (LMP) agarose gel electrophoresis. Run the sample on a 20-cm long 1% agarose gel at 30–35 V overnight at 4°C. Do not stain the DNA with EtBr and do not expose it to UV. Use EtBr-stained DNA marker lanes as a ruler to cut out the agarose region containing the 25–60-kb DNA and trim excess agarose. Then you can proceed to the agarose gel-digesting assay using the "*GELase* (*EPICENTRE*) *Agarose Gel-Digesting protocol*" described in steps 5–11 below.

10. Thoroughly melt the gel slice by incubating at 70°C for 3 min for each 200 mg of gel.
11. Transfer the molten agarose immediately to 45°C and equilibrate 2 min for each 200 mg of gel.
12. Digest the agarose with 1 U of GELase for 30 min at 45°C.
13. Centrifuge the tubes in a microcentrifuge at maximum speed (15,000 × *g*) for 15 min at 4°C to pellet any insoluble oligosaccharides. Carefully remove the upper 90–95% of the supernatant, which contains the DNA, to a sterile 1.5-mL tube. You should be careful to avoid the gelatinous pellet.
14. Then precipitate the DNA by adding one-tenth volume of 3 M sodium acetate (pH 7.0) and 2.5 volumes of ethanol and mix gently.
15. Wash the pellet with 70% ethanol. Gently resuspend the DNA pellet in TE buffer (around 200 μL).
16. Concentrate the DNA in a Microcon-100 (Millipore) concentrator membrane (100-kDa cut-off) at 4°C to a final volume of 20–50 μL. DNA concentration can be around 75 ng (in 50 μL a total of 3.75 μg). This concentrated DNA is the insert to be ligated to the pCC1FOS vector.

3.5.1.2. Ligation Reaction in the pCC1FOS Fosmid Vector

1. A single ligation reaction will produce 103–106 clones depending on the quality of the insert DNA.
2. Based on this information calculate the number of ligation reactions that you will need to perform.
3. The ligation reaction can be scaled up as needed. A 10:1 molar ratio of pCC1FOS vector to insert DNA is optimal. If we use 0.5 μg of 100-kb DNA insert, we need around 0.5 μg of vector.
4. Combine the following reagents in the order listed and mix thoroughly after each addition.

   1-μL 10× Fast-Link Ligation Buffer

   1-μL pCC1FOS (0.5 μg/μL)

   1-μL 10 mM ATP

   6.8-μL concentrated insert DNA (75 ng/μL)

   0.2-μL MilliQ water

   1-μL Fast-Link DNA Ligase

   10-μL Total reaction volume
5. Incubate at room temperature for 2 h and then transfer the reaction to 70°C for 10 min to inactivate the Fast-Link DNA Ligase, after which the construct is packaged, following the next steps.

6. Thaw, on ice, one tube of the MaxPlax Lambda Packaging Extracts for every ligation reaction performed in the above step.
7. When thawed, immediately transfer 25 μL (one-half) of each packaging extract to a second 1.5-mL microfuge tube and place on ice.
8. Add 10 μL of the ligation reaction to each 25 μL of the thawed, extracts being held on ice.
9. Mix by pipetting the solutions several times. Avoid the introduction of air bubbles. Briefly centrifuge the tubes to get all liquid to the bottom.
10. Incubate the packaging reactions at 30°C for 90 min. After the 90 min packaging reaction is complete, add the remaining 25 μL of MaxPlax Lambda Packaging Extract from to each tube.
11. Incubate the reactions for an additional 90 min at 30°C.
12. At the end of the second 90-min incubation, add Phage Dilution buffer (PD buffer) to 1-mL final volume in each tube and mix gently.
13. Add 25 μL of chloroform to each tube.
14. Mix gently and store at 4°C (up to a month). A viscous precipitate may form after addition of the chloroform. This precipitate will not interfere with library production.
15. Determine the titer of the phage particles (packaged fosmid clones) and then plate the fosmid library.
16. On the day of the packaging reactions, inoculate 50 mL of LB broth + 10 mM $MgSO_4$ with 5 mL of the EPI300-T1^R overnight culture.
17. Shake at 37°C to an OD_{600nm} = 0.8–1.0.
18. Store the cells at 4°C until needed (titering).
19. The cells may be stored for up to 72 h at 4°C, if necessary.

3.5.1.3. Titering the Packaged Fosmid Clones

Before plating the library, we recommend that the titer of packaged fosmid clones be determined. This will aid in determining the number of plates and dilutions to make to obtain a library that meets the needs of the user.

1. Make serial dilutions of the 1 mL of packaged phage particles into PD buffer in sterile microfuge tubes. For example, use dilutions $1:10^1$, $1:10^2$, $1:10^4$, and $1:10^5$.
2. Add 10 μL of each above dilution, individually, to 100 μL of the prepared EPI300-T1^R host cells. Incubate each for 20 min at 37°C.

3. Spread the infected EPI300-T1^R cells on an LB plate plus 12.5 μg/mL chloramphenicol and incubate at 37°C overnight to select for the fosmid clones.
4. Count colonies and calculate the titer of the packaged clones as following: if there are 200 colonies on the plate streaked with the 1:10^4 dilution, then the titer in cfu/mL, (where cfu represents colony-forming units) of this reaction would be:

   (No. of colonies) (dilution factor) (1,000 μL/mL)/(volume of phage plated [μL])

   That is: (200 cfu) (10^4) (1,000 μL/mL)/(10 μL) = 2×10^8 cfu/mL
5. Based on the titer of the phage particles determined before, dilute the phage particles from with PD buffer to obtain the desired number of clones and clone density on the plate.
6. Mix the diluted phage particles with EPI300-T1^R cells prepared in the ratio of 100 μL of cells (prepared as above) for every 10 μL of diluted phage particles.
7. Spread the infected bacteria on an LB plate plus 12.5 μg/mL chloramphenicol and incubate at 37°C overnight to select for the fosmid clones.
8. Subsequently, these clones are plated with the help of a colony-picker robot, in 384-well plates (LB, 12.5 μg/mL chloramphenicol and 15% of glycerol).
9. Plates are incubated overnight without shaking at 37°C.
10. The colony-picker robot is again used to produce copies of the 384-well plates.

3.5.2. Using the pLAFR3 Cosmid

Since the discovery rate of novel proteins using traditional cultivation techniques has significantly decreased during the past couple of years, many different expression hosts, apart from the usual *E. coli* systems, are used at the moment for cloning DNA fragments (28). Of particular interest is the mining and further reconstitution of natural product biosynthetic pathways where large multienzyme assemblies should be functionally expressed and where the choice of a suitable heterologous host is critical (29). In this case, it has been proposed the generation of broad host range vectors for replication in different Gram-negative species, such as pLAFR3 vector, which is able to replicate in *Pseudomonas* strains hosts (30). Strains of the *Pseudomonas* genus are known to be ubiquitous and metabolically versatile, with a great ability to metabolize toxic organic chemicals, such as aliphatic and aromatic hydrocarbons (31). This ability, along with its genetic plasticity, makes these bacteria very attractive for cloning of DNA libraries rich in biodegradation pathways. To this end, we explain in this section the protocol to generate metagenomic libraries with the pLAFR3

vector, which allow the cloning of around 23 kb insert DNA in the expression hosts of the *Pseudomonas* genus.

3.5.2.1. pLAFR3 Shoulders Preparation

1. Inoculate 200 mL of LB, Tc 10 µg/mL with a single colony of *E. coli* S17-3 (bearing pLAFR3 cosmid) and grow it overnight with orbital shaking (250 rpm) at 30°C. Pellet cells for 10 min at 7,000 × *g* and isolate pLAFR3 plasmid with large-Construct kit, treating the sample with ATP-dependent exonuclease to have just this cosmid, thus eliminating DNA chromosome.
2. Then, take two aliquots of around 3 µg of pLAFR3 and cut one with *Hin*dIII (shoulder 1) and the other with *Eco*RI (shoulder 2) at 37°C overnight. Run small aliquots in a 0.75% agarose electrophoresis gel just to see that the digestion has worked property. Then, incubate samples at 65°C for 20 min to inactivate restriction enzymes.
3. Add 3 µL of SAP to dephosphorylate DNA, incubate for 1 h at 37°C. In order to spurn DNA shearing, avoid pipetting, just stir the tube to mix. Then, incubate samples at 65°C for 20 min to inactivate SAP.
4. Mix the pLAFR3 shoulders at 1:1 and add 400 µL of water to wash it off in Microcon-100. Concentrate it to a small volume (around 30–40 µL).
5. To a volume of 37 µL of Microcon-concentrated DNA add 5 µL of buffer 10× NEB3, 5 µL of BSA 10×, 2 µL of MilliQ water and 1 µL of *Bam*HI enzyme and digest overnight at 37°C.
6. Run small aliquots in a 0.75% agarose electrophoresis gel just to see that the fragments remain in the same size (22 kb), as before *Bam*HI-digestion.
7. Use the GeneClean Kit to inactivate *Bam*HI and to concentrate the pLAFR3 shoulders.
8. Ligate overnight at 4°C shoulders and insert DNA (genomic DNA partially digested with *Sau*3AI, see preparation in the next step).
9. Packaging with Gigapack XL and select on LB, tetracycline 10 µg/mL, XGal 40 µg/mL.

3.5.2.2. Partial *Sau*3AI Digestion of DNA Insert for pLAFR3 Cloning

In order to obtain DNA fragments of 25–50 kb partially digested with *Sau*3AI is recommended to do some pilot reactions using different amounts of enzyme.

1. Set up a series of reactions starting for example from 2 U of enzyme per 1 µg of DNA, 1, 0.5, 0.25, and 0.125 U/µg. You should choose two different restriction conditions, as in the following example:

Reaction 1: 0.14 U/μL

20 μL concentrated insert DNA (12 μg)

5 μL Ligation Buffer NEB1 (10×)

5 μL BSA (10×)

2 μL MilliQ water

18 μL *Sau*3AI 0.4 U/μL

Total reaction volume: 50 μL

Reaction 2: 0.21 U/μL

20 μL concentrated insert DNA (12 μg)

7 μL Ligation Buffer NEB1 (10×)

7 μL BSA (10×)

2-μL MilliQ water

36-μL *Sau*3AI 0.4 U/μL

Total reaction volume: 70 μL

2. Incubate for 20 min at 37°C.
3. Stop reactions by adding 65 mM 0.5 M EDTA pH 8 (1.5 μL for each 10 μL reaction volume) and heat the samples to 65°C for 15 min.
4. Then, mix both reactions and load samples on a 20-cm long preparative gel 1% agarose, run it at 30–35 V overnight at 4°C, and cut and stain the slots with the DNA marker with ethidium bromide. Do not stain the part of the gel containing your DNA for cloning.
5. Under UV light, cut out the part of the gel blocks with the DNA markers in the range of approximately 20 kbp to use them as a marker to excise the gel with environmental DNA.
6. Cut out the desired gel region (25–40-kb gel region) and trim excess agarose.
7. Then, proceed to the agarose gel digestion following the GELase protocol and concentrate DNA as described above.

3.5.2.3. Ligation in pLAR3 Cosmid

Ligate overnight at 14°C partially *Sau3A*I digested DNA and pLAFR3 shoulders in a ratio 1:2 or 1:1. The ligation volume must be as low as possible (5–10 μL). If you take 100 ng of both shoulders together, then add 50 or 100 ng of the insert (you may do two separate ligations and see what works better). It is highly recommended to run small aliquots (for example 1 μL) of all your samples after any manipulation, and after ligation. Then, package with Gigapack XL, as described in the *Packaging Protocol*:

1. Remove the appropriate number of packaging extracts from a –80°C freezer and place the extracts on dry ice.

2. Quickly thaw the packaging extract by holding the tube between your fingers until the contents of the tube just begins to thaw.
3. Add the experimental DNA immediately (1–4 μL containing 0.1–1.0 μg of ligated DNA) to the packaging extract.
4. Stir the tube with a pipet tip to mix well. Gentle pipetting is allowable, provided that air bubbles are not introduced.
5. Spin the tube quickly (for 3–5 s), if desired, to ensure that all contents are at the bottom of the tube.
6. Incubate the tube at room temperature (22°C) for 2 h.
7. Add 500 μL of SM buffer to the tube. The gelatin in SM buffer stabilizes lambda phage particles during storage.
8. Add 20 μL of chloroform and mix the contents of the tube gently.
9. Spin the tube briefly to sediment the debris.
10. The supernatant containing the phage is ready for titering. The supernatant may be stored at 4°C for up to 1 month.

3.5.2.4. Titering the Cosmid Packaging Reaction

1. Streak the bacterial glycerol stock (*E. coli* DH5α or XL1Blue) onto the LB agar plates. Incubate the plates overnight at 37°C. Do not add antibiotic to the medium in the following step. The antibiotic bind to the bacterial cell wall and inhibit the ability of the phage to infect the cell.
2. Inoculate 50 mL of LB, supplemented with 10 mM $MgSO_4$ and 0.2% (w/v) maltose, with a single colony.
3. Grow at 37°C, shaking for 4–6 h (do not grow past an OD_{600} of 1.0). Alternatively, grow overnight at 30°C, shaking at 200 rpm.
4. Pellet the bacteria at 500 × *g* for 10 min.
5. Gently resuspend the cells in half the original volume with sterile 10 mM $MgSO_4$.
6. Dilute the cells to an OD_{600} of 0.5 with sterile 10 mM $MgSO_4$. The bacteria should be used immediately following dilution.
7. Prepare a 1:10 and a 1:50 dilution of the cosmid packaging reaction in SM buffer.
8. Mix 25 μL of each dilution with 25 μL of the appropriate bacterial cells at an OD_{600} of 0.5 in a microcentrifuge tube and incubate the tube at room temperature for 30 min.
9. Add 200 μL of LB broth to each sample and incubate for 1 h at 37°C, shaking the tube gently once every 15 min. This incubation allows time for expression of the antibiotic resistance.
10. Spin the microcentrifuge tube for 1 min and resuspend the pellet in 50 μL of fresh LB broth.

11. Using a sterile spreader, plate the cells on LB agar plus 10 μg/mL tetracycline and incubate at 37°C overnight to select for the fosmid clones. Incubate the plates overnight at 37°C.
12. Count colonies and calculate the titer of the packaged phage particles as is described above.
13. Based on the titer of the phage particles, dilute the phage particles from with SM buffer to obtain the desired number of clones and clone density on the plate.
14. Mix the diluted phage particles with *E. coli* DH5α or XL1Blue cells prepared in the ratio of 100 μL of cells for every 10 μL of diluted phage particles.
15. Spread the infected bacteria on LB agar, tetracycline 10 μg/mL, XGal 40 μg/mL plates and incubate at 37°C overnight to select for the plasmid clones. Subsequently, these clones are plated with the help of a colony-picker robot, in 384-well plates (LB, tetracycline 10 μg/mL, and 15% of glycerol).
16. Plates are incubated overnight without shaking at 37°C.
17. The colony-picker robot is again used to produce copies of the 384-well plates.

3.5.3. Using the Lambda Zap® Express System

Small-insert expression libraries, especially those made in lambda phage vectors, are specially constructed for activity screens; however, in contrast with cosmid or fosmid vectors, the *Zap Express pBK vector* allows cloning of up to 15 kbp (optimal about 8.5–9.5 kbp).

3.5.3.1. Partial *Sau*3AI Digestion of DNA Insert for Cloning in Zap Express Vector

1. In order to obtain DNA fragments of about 8.5–9.5 kbp partially digested with *Sau*3AI, it is recommended to do some trial reactions using different amounts of enzyme. Set up a series of reactions starting, for example, from 0.1 to 0.04 U of enzyme per 1 μg of DNA:

   Reaction 1: 0.04 U (0.004 U/μL)

   0.683-μL DNA (585.5 ng/μL)

   1.0-μL NE buffer (10×)

   1.0-μL BSA buffer (10×)

   7.23-μL H_2O

   0.1-μL *Sau*3AI 0.4 U/μL

   Reaction 2: 0.06 U (0.006 U/μL)

   0.683-μL DNA (585.5 ng/μL)

   1.0-μL NE buffer (10×)

   1.0-μL BSA buffer (10×)

   7.17-μL H_2O

   0.15-μL *Sau*3AI 0.4 U/μL

Reaction 3: 0.1 U (0.01 U/μL)

0.683-μL DNA (585.5 ng/μL)

1.0-μL NE buffer (10×)

1.0-μL BSA buffer (10×)

7.07-μL H_2O

0.25-μL *Sau*3AI 0.4 U/μL

Reaction 4: 0.5 U (0.05 U/μL)

0.683-μL DNA (585.5 ng/μL)

1.0-μL NE buffer (10×)

1.0-μL BSA buffer (10×)

6.07-μL H_2O

1.25-μL *Sau*3AI 0.4 U/μL

Reaction 5: 1 U (0.1 U/μL)

0.683-μL DNA (585.5 ng/μL)

1.0-μL NE buffer (10×)

1.0-μL BSA buffer (10×)

4.82-μL H_2O

2.5-μL *Sau*3AI 0.4 U/μL

2. Incubate 20 min at 37°C.
3. Stop reactions by adding 65 mM EDTA pH 8 and by heating the samples at 65°C for 15 min.
4. Then, run a 20-cm long gel 1% agarose at 30–35 V overnight at 4°C and stain with ethidium bromide. Use the partial digestion conditions that result in a majority of the DNA migrating in the desired size range (5–15 kb).
5. So, for the partial digestion of the DNA, you should scale up *Sau*3AI enzyme amount for at least 2 μg DNA. The two best restriction conditions are selected and scale up, as in the following example:

(R^N 1)

20-μL DNA (11.7 μg)

5-μL Buffer NE (10×)

5-μL Buffer BSA (10×)

2-μL H_2O MilliQ

18-μL *Sau*3A 0.4 U/μL (7.2 U; 0.144 U/μL R^N; 0.61 U/μg DNA)

(R^N 2)

20-μL DNA (11.7 μg)

7-μL Buffer NE (10×)

7-μL Buffer BSA (10×)

0-μL H_2O MilliQ

36-μL *Sau*3A 0.4 U/μL (14.4 U; 0.206 U/μL R^N; 1.23 U/μL DNA)

6. Incubate 20 min at 37°C.
7. Stop reactions by adding 65 mM EDTA 0.5 M pH 8 (1.5 μL for each 10 μL reaction volume) and heat the samples to 65°C for 15 min. Then, mix both reactions and load samples on a 20-cm long preparative gel 1% agarose, run it at 30–35 V overnight at 4°C and stain marker lanes with ethidium bromide.
8. Cut out the desired gel region (25–40-kb gel region) and trim excess agarose. Do not expose of the gel slice to UV light to minimize the formation of pyrimidine dimers.
9. Then, proceed to the agarose gel digestion following the GELase (EPICENTRE) protocol and concentrate DNA as described above.

3.5.3.2. Ligation in pBK-CMV Vector Predigested with *Bam*H1

1. Ligate overnight at 14°C partially *Sau*3AI digested DNA and pBK-CMV, using the following ligation conditions:

   1-μL Zap Express Vector

   0.6-μL T4 ligase buffer (10×)

   4-μL of concentrated insert

   0.6-μL T4 DNA ligase

   The final volume should not exceed 5.0–5.5 μL.
2. Further, the ligation product is packaged with Gigapack XL, as described above, and titered.
3. Streak the bacterial glycerol stock (*E. coli* XL1 MRF′) onto the LB agar plates. Incubate the plates overnight at 37°C. Do not add antibiotic to the medium in the following step. The antibiotic binds to the bacterial cell wall and inhibits the ability of the phage to infect the cell.
4. Inoculate 50 mL of LB, supplemented with 10 mM $MgSO_4$ and 0.2% (w/v) maltose, with a single colony.
5. Grow at 37°C, shaking for 4–6 h (do not grow past an OD_{600} of 1.0). Alternatively, grow overnight at 30°C, shaking at 200 rpm.
6. Pellet the bacteria at 500 × *g* for 10 min.
7. Gently resuspend the cells in half the original volume with sterile 10 mM $MgSO_4$.
8. Dilute the cells to an OD_{600} of 0.5 with sterile 10 mM $MgSO_4$. The bacteria should be used immediately following dilution.

9. Prepare dilutions from 1:1 to 1:10^5 of the packaging reaction in SM buffer.
10. Mix 1 μL of each dilution with 200 μL of the appropriate bacterial cells at an OD_{600} of 0.5 in a microcentrifuge tube and incubate the tube at 37°C for 15 min shaking the tube gently.
11. Add 500 μL of NZY soft agar to each sample plate on NZY agar plates. Incubate the plates overnight at 37°C.
12. Count phage particles and calculate the titer of the packaged phage particles, as described above.
13. After the titer, used to calculate the library size, the library is further amplified. Amplification can be performed both in liquid medium or agar plates. For amplification in liquid culture use the following steps.
14. Mix 2 mL of a fresh, overnight bacterial culture (OD_{600} 0.95) with approximately 10^6 pfu of bacteriophage in a sterile culture tube.
15. Incubate for 15 min at 37°C to allow the bacteriophage particles to adsorb.
16. Add 8 mL of prewarmed LB medium (or NZY) and incubate with vigorous shaking until lysis occurs (6–12 h at 37°C).
17. After lysis has occurred, add two drops of chloroform and continue incubation for 15 min at 37°C.
18. Centrifuge at 4,000 × *g* for 10 min at 4°C.
19. Recover the supernatant, add one drop of chloroform, and store at 4°C. The titer of the stock should be approximately 10^{10} pfu/mL, and this usually remains unchanged as long as the stock is stored at 4°C.
20. For the amplification in solid agar, *E. coli* XL1 MRF′ cells are prepared as described above in $MgSO_4$ 10 mM and OD_{600} of 0.5. Then, proceed as follows.
21. Two aliquots are prepared, each of them containing approximately 5×10^4 pfu and 600-μL *E. coli* cells. Do not exceed 300-μL phage solution per 600 μL of cells.
22. Incubate for 15 min at 37°C with gentle shaking after which 3 mL of NZY broth is added and further spread over NZY agar plate (20 cm × 20 cm) prewarmed at 37°C.
23. Incubate the plates at 37°C for about 8–10 h after which 8–10-mL of SM buffer is added while shaking the plates gently (50 rpm) for an additional 10 h at 4°C.
24. The buffer is then decanted in a Falcon tube. Two additional mL of SM buffer are added to the agar and mixed with the previous solution.

25. Add 5% (v/v) chloroform and incubate for 15 min at 4°C.
26. Centrifuge at 500 × *g* for 10 min at 4°C.
27. The supernatant is collected and stored: one small aliquot at 4°C for lab use, and other is stored at –70°C after addition of 7% DMSO. The library is then ready to use.

3.5.4. Protocol for Preparation and Digestion of Bacterial Chromosomal DNA and Cloning into pBACs

The full protocol takes up to 9 days, each of them consisting on the following steps.

3.5.4.1. Making Gel Plugs with Entrapped Cells and Cell Digestion for Preparation of DNA in Gel Plugs

This protocol results in approximately 200 plugs of the dimensions 2 mm × 5 mm × 10 mm (100 μL).

1. Grow *E. coli* to mid-log phase (~1×10^8 cells/mL) in 2× LB broth at 37°C. Then, add chloramphenicol (180-μg/mL final concentration) for chromosomal alignment and monitor growth (A_{550nm}) during the next 30–60 min of incubation until the A_{550nm} plateaus.
2. Chill cells on ice by swirling.
3. Collect cells by centrifugation at 2,500 × *g* for 10 min at 4°C.
4. Wash the cell pellet once with Pett IV buffer.
5. Resuspend the cell pellet in 10.4-mL Pett IV buffer (104.0-μg DNA/10.4 mL = 10.0 μg/mL = 1,000 ng/100 μL).
6. Warm cells in Pett IV buffer to 37°C.
7. Dilute with an equal volume of liquid 1% InCert Agarose.
8. Mix and hold at 37°C.
9. Pipette 100-μL aliquots of the cell–agarose mixture into a plug-forming mold.
10. Cool on ice for 10 min to gel. Be certain that the mold is sitting on a bed of ice to ensure adequate cooling.
11. Push gel plugs from the gel mold into a sterile screw-cap centrifuge containing two volumes of EC lysis buffer to one volume of gel plugs for digestion of cell walls, cell membranes, and cellular RNA.
12. Incubate at 37°C for 16 h with gentle shaking.

3.5.4.2. Washing Gel Plugs

1. Transfer the plugs to ESP buffer (one volume of ESP to one volume of gel plugs) and incubate at 50°C for 24 h with gentle shaking.
2. Change buffer and incubate for another 24 h.
3. For storage of the gel plugs, add fresh ESP buffer and store at 4°C. The DNA is stable in the gel plugs and ESP buffer for approximately 1 year.

3.5.4.3. Preparation of Gel Plugs for Restriction Endonuclease Digestion

1. Incubate the plugs with gentle shaking for 2 h at 37°C in TE buffer plus PMSF (five volumes of buffer to one volume of gel plugs), to inactivate the proteinase K.
2. Change solution and incubate overnight.

3.5.4.4. Restriction Endonuclease Digestion of DNA

1. Do two 2-h washes in TE buffer at 37°C (five volumes of buffer to one volume of gel plugs).
2. Add the required endonuclease enzyme and the recommended restriction buffer. You should follow the manufacturer's recommendation for restriction buffer (it usually includes at least 100 μg/mL BSA and 1-mm fresh reducing agent, either 2-mercaptoethanol or dithiothreitol). It is recommended to do some pilot digestion assays before digestion in plugs.
3. For your particular DNA and particular enzyme, start with 20 U/μg DNA to obtain complete digestion (DNA for different species may require different levels of restriction enzymes for complete digestion). Incubate for 4–16 h at the recommended temperature.
4. After obtaining a complete digestion, you may be able to titrate to a lower ratio of units of enzyme to μg DNA. For instance, complete digestion *E. coli* in InCert® agarose gel plugs can be obtained with levels ≤10 U/μg DNA.
5. Stop digestion by aspirating the restriction buffer and adding 500 μL ES buffer. Incubate at 50°C for 2 h. Remove ES buffer and add 500 μL ESP buffer.
6. The DNA is now ready for electrophoresis. The digested DNA can be stored at least 1 month in the gel plugs and ESP buffer at 4°C.

3.5.4.5. Casting the Running Gel and Loading the Gel Plug Slice onto the Running Gel

1. Slice the gel plug using a glass cover slip.
2. Load a slice of the gel plug. In the case of the *E. coli* DNA, load one-fourth of the gel plug which is approximately 125 ng DNA.
3. An alcohol/flame-sterilized glass rod is used to gently push the gel slice into the well, keeping the slice intact, until it touches the bottom of the well. Avoid trapping air bubbles, as well as mashing the gel slice, as DNA mobility may be affected. Overlay the gel slice with liquid agarose of the same type, concentration, and buffer as the running gel.
4. Cut a slice of a gel plug, the size necessary to achieve the DNA concentration required for the particular running gel and electrophoresis apparatus. The slice is then applied to the tooth of a comb with the slice placed level with the bottom of the comb. Once all of the sample slices are placed on the individual teeth of the comb, the comb is placed in the gel chamber with the slices facing the anode (in the direction of

migration). Both the comb teeth and the gel slices must touch the gel tray. Liquefied agarose is poured into the gel chamber, and once the gel forms, the comb is pulled from the gel. The gel slices remain embedded in the agarose gel. The wells are filled in with liquid agarose. Once the wells gel, add buffer and electrophorese as usual. It is recommended to run DNA standards (lambda ladders and *Saccharomyces cerevisaiae* chromosomal DNA) prepared in InCert® agarose gel plugs alongside the *E. coli* digests.

5. Load the gel into the electrophoresis until filled with PFG-TBE buffer. Set the voltage, pulse time, run time and ramp factor (if applicable) according to the instrument manufacturer's recommendations to achieve optimal separation of the DNA fragments. We routinely run *E. coli* digests on 20 cm × 20 cm, 1% SeaKem® GTG or SeaPlaque® GTG agarose gels for 40 h at 25-s pulses in a Pulsaphor® apparatus at 10 V/cm at 12°C.

3.5.4.6. Staining the Running Gel

1. Stain the gelin 1 μg/mL ethidium bromide for 20 min, followed by a 30–60-min rinse in PFG-TBE running buffer. A complete digestion of DNA is determined by sharp band patterns specific to each restriction endonuclease digestion. To minimize nicking of the DNA fragments, keep the gel in a light-opaque box when ethidium bromide is present.
2. Photograph the gel using fast film (e.g., Polaroid® type 57) to avoid lengthy exposures to UV light.

3.5.4.7. BAC Library Construction

For construction of metagenomic libraries in bacterial artificial chromosomes (BACs), vectors such as pBeloBAC11, pIndigoBAC, and pIndigoBAC-5, are recommended. They are characterized by a strict control of number of copies per cell (one or two copies), thus making cloning of DNA fragments up to 100–200 kbp more stable than plasmids, cosmids, or fosmids. These vectors also allow selection by *lacZ* promoter.

1. The DNA fragment of desired size (digested with *Bam*H1), purified as described above (see *Day 8*), is ligated overnight at 14°C with the predigested *Sau*3A-pBAC vector, using the following ligation conditions:

   1-μL 10× Ligation Buffer

   1-μL pBeloBAC11 (1.0 μg/μL)

   1-μL 10 mM ATP

   6.8-μL concentrated insert DNA (150 ng/μL)

   0.2-μL MilliQ water

   1-μL DNA Ligase

   10-μL Total reaction volume

2. The ligation products are purified by butanol precipitation using the following steps.
3. Add 50 μL MilliQ water and then 500 μL 1-butanol (there will be two phases), and vortex.
4. Centrifuge 10 min at approximately 15,000 × *g*.
5. Remove the supernatant with a pipette taking care of removing the pellet.
6. Dry the pellet in a speed-vac at room temperature.
7. Resuspend the DNA in 5 μL MilliQ. This solution can be stored at 4°C or –20°C until use.
8. The DNA is then ready for electroporation using any Epicentre IndigoBac kits (avoid heat shock). Once the colonies are obtained, they are plated by standard conditions and subsequently these clones are plated with the help of a colony-picker robot, in 384-well plates (LB plus appropriate antibiotics and 15% of glycerol). Plates are incubated overnight without shaking at 37°C. The colony-picker robot is again used to produce copies of the 384-well plates.

3.6. Activity Screens

Metagenomic libraries prepared from total environmental DNA provide comprehensive sequence information for the most abundant members of microbial community. These libraries allow for culture-independent screening for novel enzymes catalyzing the biotransformation and biodegradation of a variety of important compounds (2). Among key steps for accessing these activities, extraction of high-quality DNA from environmental samples remains a challenge due to biases in cell recovery and cell lysis. Total DNA extracted from environmental samples, in many cases, does not contain even representation of the population's genome meaning that rare organisms would contribute less to the overall DNA diversity, with the library being dominated by the most abundant organisms. This requires adapting DNA extraction methods and cloning strategies for normalization of the sample (30). Further, DNA harvested from environmental samples can be cloned into plasmids, BACs, cosmids, or fosmids for propagation in a host organism, such as *E. coli*, and subsequently screened for catalytic activity (2). Currently, *E. coli* is the most common host organism used for screening metagenomic libraries, although, in many cases, the *E. coli* transcription–translation machinery is not compatible with the expression of genes harvested from environmental microbes. This can result in a very low proportion of positive clones being obtained from one round of screening of metagenomic libraries (in some cases less than 0.01%) (32). As such, other organisms are being used for metagenomic library screening, including strains of *Streptomyces* and *Pseudomonas* (31). Several parameters are important for successful screening of

metagenomic libraries, such as the abundance of the gene in the library, the average insert size and the gene or operon length, and the use of an adequate host organism that is able to express the target gene, and other factors in *trans* to facilitate expression and folding (such as chaperones, cofactors, etc.) (31). Although these factors are not critical in homology-based screening (32), they constitute drawbacks for functional screening. In this context, the low proportion of positive clones highlights the need for sample enrichment prior to the construction of metagenomic libraries and for the use of HTS screening methodologies (including expression-independent screenings) to increase the chance of isolating novel enzymes (see below).

Independently of the organism for cloning and expression of metagenomic DNA, a function-based screening for a particular conversion, seems to be the best option to look for new proteins that do not share any homology with other known counterparts. This is of special interest as we know that about 3×10^{30} individual enzymes at species level, distributed among 10,000–15,000 protein families, are suggested to be undiscovered (2). It should also be considered that proteins and/or enzymes belonging to a superfamily are evolutionary related and share more than 50% of sequence similarity each with other. The sequence similarity suggests common structural features and often functional similarities. In this respect, potential industrial processes could be far more diverse than one can imagine simply counting all existing protein families. This, on the other hand, may offer the chance to discover many novel industrial bioconversions that are not amenable to the existing biocatalysts.

Numerous assays enable detection of enzymatic activities in agar colonies or crude cell lysates by the production of a fluorophore or chromophore [see examples in (33)]. Assays on agar-plated colonies typically enable the screening of $>10^4$ variants in a matter of days but are often limited in sensitivity. The range of assays that are applicable for crude cell lysates is obviously much wider, but their throughput is rather restricted to only 10^3–10^4 clones (34). These low-to-medium throughput screens have proved effective for the isolation of enzymes from natural or preenriched metagenomic libraries, some of which has been successfully applied at industrial scale [for extensive review see (2)].

The major enzyme activities subjected to metagenomic discovery by these methods are hydrolases, including esterases, lipases, glycosyl hydrolases, and lactonases followed by oxidoreductases, including mono- and dioxygenases. Both enzyme superfamilies constitute two of the major enzyme representatives at industrial scale. For example, recent work by Xu et al. (35) have reported the terminal oxygenation of alkanes by strains of *P. fluorescens* and *P. putida* overproducing a few alkane hydroxylases from deep-sea sediments. A prominent example of the versatility

of metagenome-based gene factories was the ability of *E. coli* to cleave a wide range of aromatic compounds (from catechol to 2,3-dihydroxybiphenyl) when expressing 38 extradiol dioxygenases retrieved from coke plant wastewater (36). Another biochemical route that starts from L-cysteine and utilizes cheaply accessible decarboxylase from alkaline polluted soils has been developed by Jiang and Wu (37) to generate the therapeutically important drug, cysteamine. More recently, the group of Jaeger has constructed a broad range of different *E. coli* strains for expressing industrially important enzymes such as benzaldehyde lyases, benzoylformate decarboxylases, hydroxynitrile lyases, and alcohol dehydrogenases, which all produced benzaldehyde by conversion of benzoin, benzoylformate, mandelonitrile, or benzoyl alcohol, respectively (38). Although these studies exemplified the potential of metagenome mining, some additional efforts should be undertaken to deliver metagenomic enzymes to the market, either as individual enzymes or as whole-cell catalysts. Till now, an esterase from uncultured microorganisms able to degrade terephtalate esters, important component of bio-plastics, is the only known example thus far (39) introduced at industrial scale, with, hopefully, many more to follow very soon. Other screening technologies use the "fluorescence-activated cell sorting" (or FACS), a technology that enables the identification of biological activity within a single cell. This system incorporates a laser with multiple wavelength capabilities and the ability to screen up to 50,000 clones per second, or over one billion clones per day. Moreover, with microtiter formats from 96 to 1,536 wells, around 10^5–10^6 assay events may be completed per day. FACS has been adapted to overcome 10^9 events per day with GigaMatrix technology that incorporates over 100,000 wells in a microtiter-sized footprint (fluorescence detection via imaging by a cooled CCD camera) (40). The substrate-dependent gene-induction assays (SIGEX) where metagenome fragments are ligated into an operon-trap vector (e.g., p18GFP) and the cells are then separated and analyzed by high-throughput FACS to select GFP-expressing cells constitute also an efficient screening tool. The mentioned SIGEX procedure requires approximately 17 days to analyze 152,000 clones with an average insertion size of 7 kbp (one billion bp). However, it requires a liquid culture by transforming a cloning host (e.g., *E. coli*) (41), thus being limited to the correct transcription and translation efficiencies and the activity of the resulting protein. Examples of enzymes analyzed by this technique are mainly enzymes involved in the degradation of aromatic compounds (41). Alternatively, the use of microarrays to profile libraries offers also an effective approach for characterizing many clones rapidly (42). This format is referred to as a metagenome microarray (MGA). In the MGA format, the "*probe*" and "*target*" concept is a reversal of those of general cDNA, and oligonucleotide

microarrays: targets (fosmid clones) are spotted on a slide and a specific gene probe is labeled and used for hybridization. This format of microarray may offer an effective approach for identifying clones from metagenome libraries rapidly without the need of laborious procedures for screening various target genes. As an example, Park et al. (42) used microarray platforms to screen microbial genomes and whole community genomes, in particular, new classes of hydrolases. However, the difficulty and limitation of this approach is related to achieving high hybridization efficiency and that the target genes derived from conserved regions of already known protein families reduce our chances for obtaining fundamentally new proteins (32).

3.6.1. Esterase Screens

1. Incubate plates containing individual colonies for 12 h at 37°C.
2. Then, cover the plates with a second layer containing the substrate [20 mL 50 mM HEPES buffer, 0.4% (w/v) agarose, 320 μL of Fast Blue RR solution in DMSO (80 mg/mL), and 320 -μL of α-naphtyl acetate solution in acetone (20 mg/mL)].
3. Positive clones appear due to the formation of a brown precipitate.

3.6.2. Cellulase-Like Screens

1. Incubate plates containing individual colonies for 12 h at 37°C in an appropriate solid medium containing 0.5–1.0% (w/v) substrate.
2. The plates are subsequently stained for 20 min with 0.5% Congo red.
3. Light zones formed with Congo red appear only if the substrate is largely hydrolyzed to oligomers with fewer than five residues.

3.6.3. P450 Oxidoreductase Screens

1. Incubate plates containing individual colonies for 12 h at 37°C.
2. Cover with a second layer containing the substrate [100 mL 50 mM Tris–HCl pH 7.5, 0.4% agarose, 50-μL of *p*-nitrophenoxycarboxylic acids such as 12-*p*NC_{10} in DMSO (15 mM)].
3. Positive clones appear due to the formation of a yellow color.

3.6.4. Laccase-Like Screens

1. Laccase production by library clones is screened by plating hybrid phage-infected cells or clones on appropriate soft agar containing 50 μM SGZ.
2. Positive clones are identified by a purple halo, produced by the oxidation of SGZ, on agar plate.

3.6.5. Screens for Sugar Fermenting Enzymes Able to Produce Alcohol

1. The presence of ethanol in the agar plate by the action of an active clone is determined by adding a solution of 0.4% agarose made up in a sodium phosphate buffer (0.1 M, pH 8), containing 50 μL 0.05 M 2,6-dichlorophenolindophenol and 100 μL 0.15 M NAD solution.
2. The solution is well shaken, poured over the agar plate, and allowed to solidify.
3. The plates were held at room temperature for 30 min before 5–20 mL 0.005 M 5-methyl-phenazinium methyl sulfate is spread over each plate by flooding or spraying and incubated further at 30°C for 30 min.
4. A yellow color indicates the reaction zone on a blue background.

3.6.6. Alcohol Oxidoreductase Screens

1. The screening is performed on indicator plates that contained 1,2-ethanediol, 2,3-butanediol, or a mixture of 1,2-propanediol and glycerol (as test substrates, although other alcohols may be used) and a mixture of pararosaniline and bisulfite for the detection of carbonyl compounds formed by the *E. coli* clones.
2. Briefly, indicator plates are prepared by adding 8 mL of pararosaniline (2.5 mg/mL of 95% ethanol; not autoclaved) and 100 mg of sodium bisulfite (unsterilized dry powder) to 400-mL batches of precooled (45°C) Luria agar lacking added carbohydrate.
3. Most of the dye is immediately converted to the leuco form by reaction with the bisulfite to produce a rose-colored medium that is dispensed into Petri plates.
4. Plates with colonies are then stored at room temperature, away from fumes which contain aldehydes (cigarette smoke, many plastic containers, etc.) and light, both of which promote increased background color.
5. Upon production of carbonyls from the test substrates, an intensely red Schiff base is formed.
6. Thus, colonies capable of carbonyl formation appear red on indicator medium and are surrounded by a red zone, whereas colonies failing to produce carbonyl compounds remain uncolored.

3.6.7. Catechol Dioxygenase Screen

1. Plates are incubated for 12 h at 37°C and are then covered with a second layer of 100 mL 50 mM Tris–HCl containing the substrate (5 μL of catechol; 0.5 mM final) and 0.4% (w/v) agarose.
2. The reaction plates are incubated at 25°C.
3. Positive clones are normally identified after incubation for 1 h for 16 h, by development of intensive yellow color.

3.6.8. Polyol Oxidase-Like Screens

1. Screening of enzymes able to oxidize methyl-β-D-galactopyranoside (or other polyol such as 2-butanol, to cite some) is performed in a 0.4% (w/v) agarose solution prepared in sodium phosphate solution that contained substrate (300 mM), catalase (700 U), and $CuSO_4$ (0.5 mM).
2. The screening is performed at room temperature and detection of a dark blue color indicates a positive clone.

4. Notes

1. Lessons from 15 years of metagenomics have taught that microbial communities are extremely complex and the community genome coverage is determining the success of a metagenomic study.
2. Above, we have described different methods to extract environmental DNA and construct environmental gene libraries. Nevertheless, the analysis of environmental samples and their comparison will never be trivial and will require a number of parallel strategies.
3. In order to compare metagenomic data from different datasets, it should also be important to set the standards for performing metagenomics projects, from physical-chemical description of sampling sites and sampling procedures down to the data interpretation and integration.

Acknowledgments

This research was supported by the BIO2006-11738, CSD2007-00005, GEN2006-27750-C-4-E, BFU2008-04398-E/BMC, and KBBE-226977 projects. A.B thanks the Spanish MEC for a FPU fellowship. P.N.G. was supported by Grant 0313751K from the Federal Ministry for Science and Education (BMBF) within the GenoMik-Plus initiative.

References

1. McHardy, A.C. and Rigoutsos, I. (2007) What's in the mix: phylogenetic classification of metagenome sequence samples. *Curr Opin Microbiol* 10, 499–503.
2. Beloqui, A., de Maria, P.D., Golyshin, P.N., and Ferrer, M. (2008) Recent trends in industrial microbiology. *Curr Opin Microbiol* 11, 240–248.
3. Lozupone, C.A. and Knight, R. (2008) Species divergence and the measurement of microbial diversity. *FEMS Microbiol Rev* 32, 557–578.

4. Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450, 560–565.
5. Schloss, P.D. and Handelsman, J. (2006) Toward a census of bacteria in soil. *PLoS Comput Biol* 2, e92.
6. DeLong, E.F. (2005) Microbial community genomics in the ocean. *Nat Rev Microbiol* 3, 459–469.
7. Ingham, C.J., Sprenkels, A., Bomer, J., Molenaar, D., van den Berg, A., van Hylckama Vlieg, J.E., et al. (2007) The micro-Petri dish, a million-well growth chip for the culture and high-throughput screening of microorganisms. *Proc Natl Acad Sci USA* 104, 18217–18222.
8. Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 68, 669–685.
9. Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., et al. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428, 37–43.
10. Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., et al. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66–74.
11. Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., et al. (2008) Functional metagenomic profiling of nine biomes. *Nature* 452, 629–632.
12. Staley, J.T. and Konopka, A. (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol* 39, 321–346.
13. Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M. (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5, R245–R249.
14. Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., et al. (2005) Comparative metagenomics of microbial communities. *Science* 308, 554–557.
15. Ferrer, M., Beloqui, A., Vieites, J.M., Guazzaroni, M.E., Berger, I., and Aharoni, A. (2009) Interplay of metagenomics and in vitro compartmentalization. *Microb Biotechnol* 2, 31–39.
16. Biddle, J.F., Fitz-Gibbon, S., Schuster, S.C., Brenchley, J.E., and House, C.H. (2008) Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. *Proc Natl Acad Sci USA* 105, 10583–10588.
17. Kalyuzhnaya, M.G., Lapidus, A., Ivanova, N., Copeland, A.C., McHardy, A.C., Szeto, E., et al. (2008) High-resolution metagenomics targets specific functional types in complex microbial communities. *Nat Biotechnol* 26, 1029–1034.
18. Bench, S.R., Hanson, T.E., Williamson, K.E., Ghosh, D., Radosovich, M., Wang, K., et al. (2007) Metagenomic characterization of Chesapeake Bay virioplankton. *Appl Environ Microbiol* 73, 7629–7641.
19. Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007) MEGAN analysis of metagenomic data. *Genome Res* 17, 377–386.
20. Lasken, R.S. (2007) Single-cell genomic sequencing using Multiple Displacement Amplification. *Curr Opin Microbiol* 10, 510–516.
21. Ishoey, T., Woyke, T., Stepanauskas, R., Novotny, M., and Lasken, R.S. (2008) Genomic sequencing of single microbial cells from environmental samples. *Curr Opin Microbiol* 11, 198–204.
22. Ferrer, M., Golyshina, O., Beloqui, A., and Golyshin, P.N. (2007) Mining enzymes from extreme environments. *Curr Opin Microbiol* 10, 207–214.
23. Bertrand, H., Poly, F., Van, V.T., Lombard, N., Nalin, R., Vogel, T.M., et al. (2005) High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction. *J Microbiol Methods* 62, 1–11.
24. Desai, C. and Madamwar, D. (2007) Extraction of inhibitor-free metagenomic DNA from polluted sediments, compatible with molecular diversity analysis using adsorption and ion-exchange treatments. *Bioresour Technol* 98, 761–768.
25. Embley, T.M. and Stackebrandt, E. (1994) The molecular phylogeny and systematics of the actinomycetes. *Annu Rev Microbiol* 48, 257–289.
26. Hugenholtz, P., Goebel, B.M., and Pace, N.R. (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* 180, 4765–4774.
27. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., et al. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239, 487–491.
28. Gabor, E.M., de Vries, E.J., and Janssen, D.B. (2004) Construction, characterization, and use of small-insert gene banks of DNA isolated

from soil and enrichment cultures for the recovery of novel amidases. *Environ Microbiol* 6, 948–958.

29. Wenzel, S.C. and Muller, R. (2005) Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Curr Opin Biotechnol* 16, 594–606.
30. Staskawicz, B., Dahlbeck, D., Keen, N., and Napoli, C. (1987) Molecular characterization of cloned avirulence genes from race 0 and race 1 of *Pseudomonas syringae* pv. *glycinea*. *J Bacteriol* 169, 5789–5794.
31. Dos Santos, V.A., Heim, S., Moore, E.R., Stratz, M., and Timmis, K.N. (2004) Insights into the genomic basis of niche specificity of *Pseudomonas putida* KT2440. *Environ Microbiol* 6, 1264–1286.
32. Gabor, E., Liebeton, K., Niehaus, F., Eck, J., and Lorenz, P. (2007) Updating the metagenomics toolbox. *Biotechnol J* 2, 201–206.
33. Andexer, J., Guterl, J.K., Pohl, M., and Eggert, T. (2006) A high-throughput screening assay for hydroxynitrile lyase activity. *Chem Commun (Camb)* 40, 4201–4203.
34. Geddie, M.L., Rowe, L.A., Alexander, O.B., and Matsumura, I. (2004) High throughput microplate screens for directed protein evolution. *Methods Enzymol* 388, 134–145.
35. Xu, M., Xiao, X., and Wang, F. (2008) Isolation and characterization of alkane hydroxylases from a metagenomic library of Pacific deep-sea sediment. *Extremophiles* 12, 255–262.
36. Suenaga, H., Ohnuki, T., and Miyazaki, K. (2007) Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environ Microbiol* 9, 2289–2297.
37. Jiang, C. and Wu, B. (2007) Molecular cloning and functional characterization of a novel decarboxylase from uncultured microorganisms. *Biochem Biophys Res Commun* 357, 421–426.
38. Henning, H., Leggewie, C., Pohl, M., Muller, M., Eggert, T., and Jaeger, K.E. (2006) Identification of novel benzoylformate decarboxylases by growth selection. *Appl Environ Microbiol* 72, 7510–7517.
39. Michels, A., Puetz, A., Maurer, K.H., Eggert, T., and Jaeger, K.E. (2007) Use of Esterases for Separating Plastics. WO 2007017181. WIPO, Geneva.
40. Robertson, D.E. and Steer, B.A. (2004) Recent progress in biocatalyst discovery and optimization. *Curr Opin Chem Biol* 8, 141–149.
41. Uchiyama, T. and Watanabe, K. (2008) Substrate-induced gene expression (SIGEX) screening of metagenome libraries. *Nat Protoc* 3, 1202–1212.
42. Park, H.J., Jeon, J.H., Kang, S.G., Lee, J.H., Lee, S.A., and Kim, H.K. (2007) Functional expression and refolding of new alkaline esterase, EM2L8 from deep-sea sediment metagenome. *Protein Expr Purif* 52, 340–347.

Chapter 2

Construction of Small-Insert and Large-Insert Metagenomic Libraries

Carola Simon and Rolf Daniel

Abstract

The vast majority of the Earth's biological diversity is hidden in uncultured and yet uncharacterized microbial genomes. The construction of metagenomic libraries is a cultivation-independent molecular approach to assess this unexplored genetic reservoir. In the last few years, a high number of novel biocatalysts have been identified by function-based or sequence-based screening of metagenomic libraries. Here, we describe detailed protocols for the construction of metagenomic small-insert and large-insert libraries in plasmids and fosmids, respectively, from environmental DNA.

Key words: Metagenomic DNA, Small-insert library, Large-insert library, Plasmid, Fosmid, Whole genome amplification

1. Introduction

The construction and screening of metagenomic libraries that have been generated from DNA directly isolated from environmental samples have been proven to be a powerful tool for the recovery of novel biomolecules of biotechnological importance (1, 2). In principle, metagenomic libraries provide access to the entire gene content of a habitat (2). The construction of metagenomic libraries involves the same steps as the cloning of genomic DNA derived from individual microorganisms. The required steps include fragmentation of environmental DNA by restriction digestion or shearing, insertion into an appropriate vector system, and transformation of the recombinant vectors into a suitable host, which is in almost all published studies on construction of metagenomic libraries *Escherichia coli* (3). Although the generation of metagenomic libraries is conceptually simple, the community

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668, DOI 10.1007/978-1-60761-823-2_2,

sizes of most metagenomes such as those derived from soil and sediment samples and, correspondingly, the large number of clones that is necessary for a significant coverage of the metagenome are great technological challenges (4, 5). Two types of libraries with respect to average insert size can be generated: small-insert libraries in plasmid vectors (less than 10 kb) and large-insert libraries in cosmid and fosmid vectors (up to 40 kb) or BAC vectors (more than 40 kb). The selection of a vector system for library construction depends on the quality of the isolated environmental DNA, the desired average insert size of the library, the copy number required, the host, and the screening strategy that will be used (3, 5). Environmental DNA that is contaminated with humic or matrix substances after purification or DNA sheared during purification is only suitable for the generation of small-insert libraries (3). Small-insert metagenomic libraries are useful for the isolation of single genes or small operons encoding novel biomolecules. To identify complex pathways encoded by large gene clusters or large DNA fragments for the partial genomic characterization of uncultured microorganisms, the generation of large-insert libraries is the appropriate method. Here, we describe one protocol for the construction of small-insert libraries and one for large-insert fosmid libraries. Both methods have been proven to be suitable for cloning of DNA purified from various environmental samples, including soil, ice, and compost (6–8).

2. Materials

2.1. Metagenomic DNA

The construction of metagenomic libraries derived from environmental samples and cloning of functional genes is dependent on the high quality of the extracted DNA, since the enzymatic modifications required during the construction of the libraries are sensitive to contamination by various biotic and abiotic components. High molecular environmental DNA is especially required for the construction of large-insert libraries. To start with library construction 5–10 μg of purified environmental DNA are required.

2.2. Generation of Small-Insert Metagenomic Libraries

1. Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare, Munich, Germany).
2. Phi29 DNA polymerase (10 U/μL) and reaction buffer (10×) (Fermentas, St. Leon-Rot, Germany).
3. S1 nuclease (100 U/μL) and reaction buffer (5×) (Fermentas, St. Leon-Rot, Germany).
4. DNA polymerase I (10 U/μL) and reaction buffer (10×) (Fermentas, St. Leon-Rot, Germany).
5. Nebulizer (Invitrogen, Karlsruhe, Germany).

6. Shearing buffer: 10 mM Tris–HCl, pH 7.5, 1 mM EDTA, 10% (w/v) glycerol. Store at room temperature.
7. Low melting point (LMP) Biozym Plaque *GeneticPure* Agarose (Biozym Scientific GmbH, Hessisch Oldendorf, Germany).
8. Tris-acetate-ethylenediamine tetraacetic acid (TAE) buffer (50×): 242 g Tris-base, 57.1 mL acetic acid, 100 mL 0.5 M EDTA, pH 8. Add H_2O to 1 L. Store at room temperature.
9. GELase Agarose Gel-Digesting Preparation (EPICENTRE Biotechnologies, Madison, WI).
10. 3 M sodium acetate, pH 5.
11. 5 M NH_4OAc, pH 7.
12. T4 DNA polymerase (5 U/μL) (Fermentas, St. Leon-Rot, Germany).
13. 10 mM dNTP Mix (Fermentas, St. Leon-Rot, Germany).
14. Klenow Fragment (10 U/μL) (Fermentas, St. Leon-Rot, Germany).
15. Buffer O (10×) (Fermentas, St. Leon-Rot, Germany).
16. SureClean (Bioline, Luckenwalde, Germany).
17. *Taq* DNA polymerase and reaction buffer with $(NH_4)_2SO_4$ (10×) (Fermentas, St. Leon-Rot, Germany).
18. 25 mM $MgCl_2$.
19. 100 mM dATP.
20. Antarctic phosphatase and buffer (10×) (New England Biolabs, Ipswich, MA).
21. Topo® XL PCR Cloning Kit (Invitrogen, Karlsruhe, Germany).
22. Bio-Rad Gene Pulser II (Bio-Rad, Munich, Germany).
23. Kanamycin stock solution: 25 mg/mL H_2O. Filter-sterilize and store at −20°C.
24. Isopropyl-β-d-thiogalactopyranoside (IPTG) stock solution: 24 mg/mL in H_2O. Filter-sterilize, divide into 2 mL aliquots and store at −20°C.
25. 5-Bromo-4-chloro-3-indolyl-β-d-galactoside (X-gal) stock solution: 20 mg/mL *N,N'*-dimethyl formamide. Filter-sterilize and store at −20°C.
26. Luria-Bertani (LB) agar: 10 g NaCl, 10 g tryptone, 5 g yeast extract per liter, pH 7.2. Add 1.5% agar. Sterilize by autoclaving.
27. LB agar supplemented with 50 μg/mL kanamycin, 48 μg/mL IPTG, and 40 μg/mL X-gal; add 1 mL of kanamycin, IPTG, and X-gal stock solution to 500 mL hot liquid LB agar after autoclaving.

2.3. Generation of Large-Insert Metagenomic Libraries

1. CopyControl™ Fosmid Library Production Kit (EPICENTRE Biotechnologies, Madison, WI). Store according to manufacturer's instructions.
2. LMP Biozym Plaque *GeneticPure* Agarose (Biozym Scientific GmbH, Hessisch Oldendorf, Germany).
3. Biometra Rotaphor (Biometra, Goettingen, Germany).
4. Tris–borate–EDTA (TBE) buffer (5×): 54 g Tris-base, 27.5 g boric acid, 20 mL 0.5 M EDTA, pH 8. Add H_2O to 1 L. Store at room temperature.
5. SureClean (Bioline, Luckenwalde, Germany).
6. LB broth supplemented with 10 mM $MgSO_4$.
7. Chloramphenicol stock solution: 6.25 mg/mL ethanol. Store at −20°C.
8. LB agar supplemented with 12.5 μg/mL chloramphenicol; add 1 mL of chloramphenicol stock solution to 500 mL molten agar.
9. 3 M sodium acetate, pH 7. Store at room temperature.
10. Phage dilution buffer: 10 mM Tris–HCl, pH 8.3, 100 mM NaCl, 10 mM $MgCl_2$. Store at room temperature.

3. Methods

Library reconstruction comprises several separate steps. For successful cloning of environmental DNA, it is recommended to avoid storage of the isolated DNA for longer periods between the individual steps. If this is not applicable, the purified DNA can be stored 1 to several days at 4°C after each step. Before conducting the end-repair of insert DNA for construction of the plasmid library (see Subheading 3.1.5) or the size fractionation for fosmid library construction (see Subheading 3.2.2), the DNA can be stored at −20°C. However, after end-repair or size fractionation, the DNA should not be stored at −20°C, as freezing and thawing will break the DNA strands. Similarly, unnecessary pipetting of the prepared DNA should be avoided. Where possible, the reagents should be added to the DNA rather than transferring the DNA. When DNA has to be transferred to a fresh microcentrifuge tube, use only large bore or cut off pipette tips to avoid further shearing of the DNA.

After completion of each step, the DNA concentration should be measured to ensure that a sufficiently high DNA concentration is recovered to conduct the remaining steps. Preferably, a large amount of DNA should be used to start as performing the separate procedures will result in the loss of DNA. If less than 5 μg of

environmental DNA are available, for reconstruction of small-insert libraries the amount of DNA can be increased by employing whole genome amplification (WGA). To improve cloning efficiency and to avoid abnormal insert size distribution, hyperbranched structures generated during WGA are resolved as described recently (9) with modifications.

In Subheadings 3.1.1–3.1.3, a protocol for WGA of the environmental DNA and resolving hyperbranched structures is given. However, if a sufficient amount of environmental DNA is available, metagenomic library construction starts with Subheading 3.1.4.

3.1. Generation of Small-Insert Metagenomic Libraries

3.1.1. Whole Genome Amplification of Environmental DNA

1. Conduct WGA of environmental DNA by using, e.g., the Illustra GenomiPhi V2 DNA Amplification Kit according to manufacturer's instructions (10).
2. Purify the DNA with SureClean according to manufacturer's instructions (11). Do not air-dry the pellet for longer than 5–10 min.
3. Resuspend the DNA pellet in 30 μL H_2O (see Note 1).

3.1.2. Resolving Hyperbranched DNA Structures (See Note 2)

1. Combine the following ingredients in a sterile microcentrifuge tube: the amplified and purified DNA from Subheading 3.1.1, step 3, 5 μL 10 mM dNTP Mix, 5 μL phi29 buffer (10×), and 1 μL phi29 DNA polymerase (10 U/μL). Add up to a final volume of 50 μL with H_2O. The reaction mix can be scaled up as needed.
2. Incubate at 30°C for 2 h.
3. Inactivate the enzyme at 65°C for 3 min.
4. Purify the DNA with SureClean (see Subheading 3.1.1, steps 2 and 3).

3.1.3. S1 Nuclease Treatment (See Note 2)

1. Set up the reaction mix as follows: the purified DNA from Subheading 3.1.2, step 4, 10 μL S1 nuclease buffer (5×), 0.5 μL S1 nuclease (100 U/μL). Add up to a final volume of 50 μL with H_2O.
2. Incubate at 37°C for 30 min.
3. Purify the DNA with SureClean (see Subheading 3.1.1, steps 2 and 3).

3.1.4. Shearing of Metagenomic DNA

1. Test the proportion of sheared DNA by running 1–2 μL of the DNA solution on a 0.8% agarose gel. If more than 50% of the DNA fragments display the desired insert size proceed with Subheading 3.1.5.
2. Assemble the nebulizer as indicated by the manufacturer.
3. Add 10 μg environmental DNA to 750 μL of shearing buffer and transfer into the bottom of the nebulizer (see Note 3).

4. Screw on cap of the nebulizer and place on ice to keep the DNA cold.
5. Connect the nebulizer to the compressed gas or air source and shear the DNA by applying 9–10 psi for approximately 10–15 s to obtain DNA fragments that are 3–8 kb in size. Check the DNA on a 0.8% agarose gel to ensure that more than 50% of the DNA fragments display the desired insert size. To vary the size of the DNA fragments either change the applied pressure or vary the time for shearing.
6. Transfer the DNA to two sterile microcentrifuge tubes.
7. Precipitate DNA by adding one of ten volume of 3 M sodium acetate, pH 5, and 2.5 volumes of 96% ethanol. Mix gently. Leave the DNA on ice for 20 min, then centrifuge in a microcentrifuge at top speed for 30 min at 4°C.
8. Discard supernatant. Subsequently, wash the pellet twice with cold 70% ethanol. After the second washing step, carefully invert the tube and allow the pellet to air-dry for 5–10 min.
9. Gently resuspend the DNA in 36 μL H_2O.

3.1.5. End-Repair of Insert DNA

1. Add the following reagents to the resuspended DNA from Subheading 3.1.4, step 9: 5 μL Buffer O (10×), 1 μL 10 mM dNTP Mix, 1 μL T4 DNA polymerase (5 U/μL), and 1 μL DNA polymerase I (10 U/μL). Add H_2O to a final volume of 50 μL (see Note 4).
2. Incubate the reaction mixture for 3 h at room temperature.
3. Inactivate the enzymes for 10 min at 75°C.

3.1.6. Size Fractionation of the Insert DNA

1. Run the blunt-ended DNA on a 1% LMP agarose gel prepared with 1× TAE buffer and a DNA size marker at each of the outside lanes of the gel. Do not include ethidium bromide in the gel.
2. Following electrophoresis, cut off the outer lanes of the gel containing the DNA ladder and stain with ethidium bromide. Visualize the DNA ladder with UV light and mark the position of the desired fragment sizes on both DNA ladders. After removing the gel slices from the UV light, reassemble the gel and cut out a gel slice containing DNA with the desired fragment size.
3. Weigh the gel slice in a tared tube.
4. Exchange the electrophoresis buffer in the gel slice with 1× GELase buffer by adding 3 μL of 1× GELase buffer per mg of gel. Incubate at room temperature for 1 h and subsequently remove the buffer (see Note 5).

5. Melt the LMP gel by incubation at 70°C for 3 min for each 200 mg of gel. If required, continue incubating at 70°C for a few more minutes.
6. Transfer the molten agarose to 45°C and equilibrate 2 min for each 200 mg of gel. Temperatures higher than 45°C will inactivate the GELase enzyme.
7. Add 1 U of GELase enzyme for each 600 mg of gel. Keep the digested agarose solution at 45°C and gently mix. Incubate for at least 1 h.
8. Transfer the reaction mixture to 70°C to inactivate the enzyme for 10 min.
9. Chill tube on ice for 5 min. Centrifuge in a microcentrifuge at top speed for 20 min to pellet any insoluble oligosaccharides. Carefully remove the supernatant and transfer to a new tube.
10. Precipitate the DNA by adding one volume of 5 M NH_4OAc, pH 7, to the molten agarose and 4 volumes of 96% ethanol (see Note 6). In the following, proceed as described in Subheading 3.1.4, steps 7 and 8.
11. Gently resuspend the DNA in 50 μL H_2O.

3.1.7. Addition of 3′ A-Overhangs to Blunt-Ended, Size-Fractionated DNA

1. Add the following reagents to the resuspended DNA from Subheading 3.1.6, step 11: 7 μL *Taq* DNA polymerase buffer (10×), 6 μL 25 mM $MgCl_2$, 1 μL 100 mM dATP, and 1 μL *Taq* DNA polymerase (5 U/μL). Add H_2O to a final volume of 70 μL.
2. Incubate at 72°C for 30 min.
3. Purify DNA by using SureClean (see Subheading 3.1.1, step 2).
4. Resuspend DNA pellet in 30 μL H_2O (see Note 7).

3.1.8. Dephosphorylation of Insert DNA

1. Prepare a reaction mixture containing the following ingredients: 12.5 μL prepared insert DNA (approximately 500 ng), 1.5 μL Antarctic phosphatase buffer (10×), 1 μL Antarctic phosphatase (5 U/μL).
2. Incubate for 15 min at 37°C.
3. Inactivate the enzyme at 65°C for 5 min.

3.1.9. TOPO® Cloning

1. Set up the following cloning reaction in a sterile microcentrifuge tube: 4 μL dephosphorylated insert DNA and 1 μL pCR®-XL-TOPO® vector.
2. Mix gently without pipetting the solution and incubate for 5 min at room temperature.

3. Add 1 µL of the TOPO® Cloning Stop Solution (6×) and mix gently.
4. Briefly centrifuge the tube and place on ice. The ligation mix may be stored for 24 h at 4°C.
5. Add 2 µL of the cloning reaction to one vial of Invitrogen's One Shot® electrocompetent *E. coli* cells and mix gently. Do not pipet.
6. Transfer cells and DNA to a prechilled 0.1 cm electroporation cuvette.
7. Electroporate the cells. We use a Bio-Rad Gene Pulser II with the following settings: 200 Ω, 25 µF, and 2.5 kV.
8. Immediately add 450 µL of room temperature S.O.C. medium (included in the Topo® XL PCR Cloning Kit) and mix well.
9. Transfer the solution to a 15 mL tube and shake horizontally for 1 h at 37°C and 150 rpm.
10. Spread 25 µL of the suspension on LB plates containing 50 µg/mL kanamycin, 48 µg/mL IPTG, and 40 µg/mL X-gal.
11. Incubate the plates overnight at 37°C.
12. Ensure that the plasmid library contains the desired insert size. Randomly pick several *E. coli* clones, grow each overnight in 5 mL LB broth supplemented with 50 µg/mL kanamycin, extract, digest, and analyze plasmid DNA by using standard techniques.
13. Count obtained clones and determine the blue/white ratio, which indicates the amount of insert-containing plasmids.
14. Extract total plasmid DNA by using standard techniques and store at −20°C.

3.2. Generation of Large-Insert Metagenomic Libraries

3.2.1. Preparation of Host Cells

1. Streak the *E. coli* EPI300-T1® cells on a LB plate. The cells are included in the CopyControl™ Fosmid Library Production Kit. Incubate overnight at 37°C. Seal the plate and store at 4°C.
2. The day before performing the lambda packaging reaction (see Subheading 3.2.6) inoculate 5 mL of LB broth with a single colony of EPI300-T1® cells and incubate overnight at 37°C and 150 rpm.

3.2.2. Shearing of Metagenomic DNA (See Note 8)

1. Randomly shear the environmental DNA by passing it several times through a small bore pipette tip.
2. Load 1–2 µL of the DNA on an agarose gel and check if more than 50% of the DNA fragments display the desired insert size. If not, repeat step 1 until sufficiently sheared DNA is obtained.

3.2.3. Size Fractionation of the Insert DNA (See Note 9)

1. Size-select the sheared metagenomic DNA as described in Subheading 3.1.6 with the following modifications.
2. Run the DNA on a 1% LMP agarose gel prepared with 1× TBE buffer using pulsed field gel electrophoresis. We use a Biometra Rotaphor with voltage and ramp times as recommended by the manufacturer. Load 100 ng of fosmid control DNA into each of the outside lanes of the gel with the environmental DNA.
3. Heat the GELase buffer (50×) (included in the CopyControl™ Fosmid Library Production Kit) to 45°C and melt the LMP agarose by incubating the tube at 70°C for 10–15 min. Transfer the tube to 45°C.
4. Add the preheated GELase buffer (50×) to 1× final concentration. Per 100 μL of molten agarose add 1 U of GELase and gently mix. Incubate for 1 h. Proceed with steps 8–11 in Subheading 3.1.6.

3.2.4. End-Repair of Insert DNA

1. Add the following reagents, which are all included in the CopyControl™ Fosmid Library Production Kit, to the 50 μL resuspended size-fractioned DNA from Subheading 3.2.3, step 4: 8 μL end-repair buffer (10×), 8 μL 2.5 mM dNTP Mix, 8 μL 10 mM ATP, 4 μL end-repair enzyme mix. Add H_2O to a final volume of 80 μL.
2. Incubate at room temperature for 2 h.
3. Inactivate the enzyme mix at 70°C for 10 min.
4. Purify the blunt-ended DNA with SureClean (see Subheading 3.1.1, step 2).
5. Resuspend the DNA in 20–30 μL H_2O (see Note 7).

3.2.5. Ligation

1. Add the following reagents, which are also included in the CopyControl™ Fosmid Library Production Kit, to the end-repaired insert DNA (Approximately 600 ng): 1 μL Fast-Link ligation buffer (10×), 1 μL 10 mM ATP, 1 μL CopyControl™ pCC1FOS Vector (0.5 μg/μL), 1 μL Fast-Link DNA ligase (2 U/μL). Add H_2O to a final volume of 10 μL.
2. Incubate overnight at 16°C.
3. Add 0.5 μL Fast-Link DNA ligase to the reaction mixture and incubate for another 1.5 h at room temperature.
4. Stop the reaction at 70°C for 10 min.

3.2.6. Packaging of Fosmids

1. Inoculate 50 mL LB broth supplemented with 10 mM $MgSO_4$ with 5 mL of an overnight culture of the EPI300-T1® cells (see Subheading 3.2.1, step 2). Incubate the culture at 37°C and 150 rpm until an OD_{600} of 0.8–1.0. Store the cells at 4°C for up to 72 h when required.

2. Thaw one tube of the MaxPlax Lambda Packaging Extracts (included in the CopyControl™ Fosmid Library Production Kit) on ice.
3. Immediately transfer 25 µL of the packaging extract to a new microcentrifuge tube on ice. Store the remaining 25 µL of the MaxPlax Packaging Extract to –70°C until use. Do not expose the packaging extracts to CO_2 sources such as dry ice.
4. Add the ligation reaction to the thawed packaging extracts on ice. Mix the solution without producing air bubbles. Briefly centrifuge the tube.
5. Incubate the reaction mixture for 90 min at 30°C.
6. Thaw the remaining packaging extract from step 3 and add it to the reaction mixture.
7. Incubate for an additional 90 min at 30°C.
8. Add phage dilution buffer to a final volume of 1 mL and mix gently. Add 25 µL chloroform and mix gently. Store at 4°C for up to 2 days.

3.2.7. Transduction of Host Cells

1. Add 10, 20, 30, 40, and 50 µL of the packaged phage particles individually to 100 µL of the prepared EPI300-T1® cells from Subheading 3.2.6, step 1.
2. Incubate for 45 min at 37°C.
3. Spread the infected EPI300-T1® cells on an LB plate supplemented with 12.5 µg/mL chloramphenicol and incubate overnight at 37°C.
4. Count colonies and mix the remaining packaged phage particles with the host cells in the ratio, which yielded the highest amount of fosmid-containing *E. coli* clones.
5. Incubate for 45 min at 37°C.
6. Ensure that the fosmid library contains the desired insert size. For this purpose, pick randomly several *E. coli* clones, grow each in 5 mL LB broth supplemented with 12.5 µg/mL chloramphenicol overnight at 37°C and 150 rpm.
7. To induce a high copy number of the fosmids in the host cells combine 500 µL of the overnight culture from step 6, 5 µL of the CopyControl™ Induction Solution (1,000×), and 4.5 mL LB broth supplemented with 12.5 µg/mL chloramphenicol in a 15 mL tube.
8. Shake the tubes at 37°C horizontally for 5 h vigorously as aeration is critical for induction of a high copy number.
9. Extract, digest, and analyze the fosmid DNA by standard techniques to ensure that the fosmid library contains metagenomic DNA.

10. Store the fosmid library in microtiter plates containing LB broth supplemented with 12.5 μL chloramphenicol at –70°C.

4. Notes

1. If the DNA pellet is difficult to resuspend, add another 20 μL of H_2O and heat to 37°C for 30 min.
2. WGA of DNA results in a hyperbranched structure, which has to be resolved prior to cloning. By incubating the amplified DNA with phi29 polymerase without primers the density of branching junctions is reduced. Resulting 3′ single-stranded overhangs are removed by S1 nuclease treatment. Nicks in the resulting double-stranded DNA are removed by incubation with DNA polymerase I, which can be performed during end-repair of the insert DNA (see Subheading 3.1.5).
3. Shearing of the metagenomic DNA can be done either mechanically using a Nebulizer or a HydroShear® (Zinsser Analytic, Frankfurt, Germany), or by partial restriction endonuclease digestion using, e.g., Bsp143I (Fermentas, St. Leon-Rot, Germany). Note that restriction endonuclease digestion will lead to more biased libraries than mechanical shearing of DNA.
4. If environmental DNA was not subjected to WGA, instead of DNA polymerase I the Klenow fragment should be added to the reaction mixture. DNA polymerase I exhibits not only polymerase and proofreading activity, but also 5′–3′ exonuclease activity, which is important for removal of nicks, which originate from the S1 nuclease treatment described in Subheading 3.1.3.
5. Size fractionation of the insert DNA can also be done by gel extraction via columns, e.g., by using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). Gel purification via columns is less time-consuming but may result in breaking of the prepared DNA strands.
6. The oligosaccharides produced by GELase digestion are more soluble in ethanol in the presence of ammonium. When other salts are used for precipitation, coprecipitation of oligosaccharides may occur.
7. If the DNA concentration is too low after complete resuspension of the DNA pellet, the DNA solution can be concentrated by freeze-drying. We use a Savant SpeedVac Plus SC110A (Thermo Fisher Scientific, Waltham, MA).
8. In some cases, this step can be omitted as DNA extraction from environmental samples frequently results in sufficiently sheared DNA. Therefore, prior to cloning, the molecular

weight of the isolated DNA should be checked by agarose gel electrophoresis.

9. Alternatively, if only a small amount of environmental DNA is available, the size fractionation step can be omitted. Only DNA fragments of approximately 40 kb will be packaged. However, without size fractionation chimeras may form. Size fractionation of the insert DNA is recommended when large contiguous DNA fragments are needed.

References

1. Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 68, 669–685.
2. Steele, H.L., Jaeger, K.-E., Daniel, R., and Streit, W.R. (2009) Advances in recovery of novel biocatalysts from metagenomes. *J Mol Microbiol Biotechnol* 16, 25–37.
3. Daniel, R. (2005) The metagenomics of soil. *Nat Rev Microbiol* 3, 470–478.
4. Riesenfeld, C.S., Schloss, P.D., and Handelsman, J. (2004) Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet* 38, 525–552.
5. Daniel, R. (2004) The soil metagenome – a rich resource for the discovery of novel natural products. *Curr Opin Biotechnol* 15, 199–204.
6. Jogler, C., Lin, W., Meyerdierks, A., Kube, M., Katzmann, E., Flies, C., Pan, Y., Amann, R., Reinhardt, R., and Schüler, D. (2009) Towards cloning the magnetotactic metagenome: Identification of magnetosome island gene clusters in uncultivated magnetotactic bacteria from different aquatic sediments. *Appl Environ Microbiol* 75, 3972–3979.
7. Simon, C., Herath, J., Rockstroh, S., and Daniel, R. (2009) Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. *Appl Environ Microbiol* 75, 2964–2968.
8. Waschkowitz, T., Rockstroh, S., and Daniel, R. (2009) Isolation and characterization of metalloproteases with a noval domain structure by construction and screening of metagenomic libraries. *Appl Environ Microbiol* 75, 2506–2516.
9. Zhang, K., Martiny, A.C., Reppas, N.B., Barry, K.W., Malek, J., Chisholm, S.W., and Church, G.M. (2006) Sequencing genomes from single cells by polymerase cloning. *Nat Biotechnol* 24, 680–686.
10. GE Healthcare. Illustra™ GenomiPhi V2 DNA Amplification Kit: instruction manual. GE Healthcare, Munich. http://www.gehealthcare.com/usen/index.html.
11. Bioline. SureClean: instruction manual. Bioline, Luckenwalde. http://www.bioline.com/.

Chapter 3

Construction and Screening of Marine Metagenomic Libraries

Nancy Weiland, Carolin Löscher, Rebekka Metzger, and Ruth Schmitz

Abstract

Marine microbial communities are highly diverse and have evolved during extended evolutionary processes of physiological adaptations under the influence of a variety of ecological conditions and selection pressures. They harbor an enormous diversity of microbes with still unknown and probably new physiological characteristics. Besides, the surfaces of marine multicellular organisms are typically covered by a consortium of epibiotic bacteria and act as barriers, where diverse interactions between microorganisms and hosts take place. Thus, microbial diversity in the water column of the oceans and the microbial consortia on marine tissues of multicellular organisms are rich sources for isolating novel bioactive compounds and genes. Here we describe the sampling, construction of large-insert metagenomic libraries from marine habitats and exemplarily one function based screen of metagenomic clones.

Key words: Isolation of metagenomic DNA, 16S rDNA phylogenetic analysis, Fosmid library, Function-based screen

1. Introduction

Current estimates indicate that more than 99% of the microorganisms present in many natural environments are not readily culturable with conventional approaches (1). To overcome the difficulties and limitations associated with cultivation techniques, several DNA-based molecular methods have been developed to explore the diversity and potential of microbial communities (2–5). The new and rapidly developing field of so-called "metagenomics" tries to analyze the complex genomes and genomic informations of microbial communities present in the different environmental habitats. Primarily employed to study nonculturable microbiota for a better understanding of global microbial ecology in different environmental niches (6), metagenomic data also provide

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668, DOI 10.1007/978-1-60761-823-2_3,

information on the functional role of the different microbes within the community. This is emphasized by several recent examples, e.g., the discovery of a new bacterial rhodopsin, proteorhodopsin (7–9) and the recent insights into symbiosis between a marine oligochaete and its microbial community (10). In recent years, efficient DNA isolation techniques for various habitats and vector systems for cloning large metagenomic DNA fragments (such as cosmids, fosmids, or BACs) allowing to screen large clone libraries for functional activities have been established and are available as commercial kits (11, 12).

The oceans are the largest ecological systems on earth (13) harboring marine microorganisms with an average cell density of approximately 5×10^5 cells/mL, leading to the estimation that the oceans are a living space for approximately 3.6×10^{28} microorganisms (14). Marine microbial communities are highly diverse and have evolved during extended evolutionary processes of physiological adaptations under the influence of a variety of ecological conditions and selection pressures. They harbor an enormous diversity of metabolically complex microbes with still unknown and probably new physiological characteristics and are thus rich sources for isolating novel bioactive compounds and genes (15). Microbes are also known to form symbiotic relationships with various marine invertebrates, e.g., sponges, corals, and squids, and are thus suspected to produce particular biologically active and pharmacologically valuable natural products (16). Furthermore, the microbial consortia on marine multicellular organisms are attractive model systems to understand the complex interplay between microbes and their host cells that may be also relevant to the human barrier organs and its microbiota providing insight into the development of human diseases and identification of new drug targets.

2. Materials

2.1. Sampling

2.1.1. Marine Water Sampling

1. Membrane pump with respective membranes (polycarbonate or polyvinylidenfluoride membrane filters of 10 and 0.22 μm pore size) or a conductivity, temperature, depth sensor (CTD) equipped with a 24 Niskin 10 L bottle rosette.
2. Peristaltic pump to accelerate the filtration.
3. In situ pumps for marine deep water sampling.
4. Liquid nitrogen to freeze the filters for long-term storage at –80°C.

2.1.2. Sampling from Marine Invertebrates

1. Equipment for sampling marine organisms, e.g., clean buckets, bottles, a dip net.

2. Autoclaved seawater to wash away loosely attached microorganisms.
3. Sterile Petri dishes and sterile cotton-tipped applicators to swab microorganisms from the surfaces of the marine eukaryote.
4. Liquid nitrogen to freeze the filters for long-term storage at −80°C.

2.2. Isolation of Metagenomic DNA

1. 37 and 65°C incubator, centrifuge.
2. DNA extraction buffer: 100 mM Tris–HCl pH 8.0, 100 mM sodium–EDTA, 100 mM sodium phosphate, 1.5 M NaCl, 1% cetyl trimethylammonium bromide (CTAB) (vol/vol).
3. TE buffer: 10 mM Tris-HCl pH 8.0, 1 mM EDTA.
4. 20 mg/mL Proteinase K (Fermentas, St. Leon-Rot), 50 mg/mL Lysozyme (Roth, Karlsruhe), RNase A (Qiagen, Hilden), 20% sodium dodecyl sulfate (SDS), chloroform, 100% isopropanol, 70% ethanol.

2.3. 16S rDNA Phylogenetic Analysis

1. Reaction tubes, pipettes, thermocycler.
2. Bacteria-specific primer 27F (5′-AGAGTTTGATCCTGGCTCAG-3′) and the universal primer 1492R (5′-GGTTACCTTGTTACGACTT-3′)
3. Archaea-specific primer set 20F (5′-TTCCGGTTGATCCCTGCCAGG-3′) and 958R (5′-TCCGGCGTTGAACTCCAATT-3′).
4. 10× *Taq* reaction buffer, 25 mM $MgCl_2$, 10 mM dNTPs, *Taq* polymerase (e.g., Fermentas, St. Leon-Rot), sterile water.
5. Gel Extraction and Purification Kit (e.g., Macherey-Nagel, Düren).
6. TA Cloning Kit (e.g., Invitrogen, Karlsruhe).
7. Sequencing Reaction Kit (BigDye® Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems, Darmstadt) and a capillary sequencer.

2.4. Construction of a Metagenomic Large-Insert Library

1. CopyControl™ Fosmid Library Production Kit (Epicentre, Madision, WI).
2. TE buffer: 10 mM Tris-HCl pH 8.0, 1 mM EDTA.
3. 0.025 μm cellulose filters type VS from Millipore (Schwalbach).
4. Phage-dilution buffer: 10 mM Tris–HCl pH 8.3, 100 mM NaCl, 10 mM $MgCl_2$.
5. LB containing 10 mM $MgSO_4$ for growth of EPI300-T1R host cells.

6. LB plates supplemented with 12.5 μg/mL chloramphenicol.
7. Microtiter plates (96 wells) containing 150 μL LB supplemented with 12.5 μg/mL chloramphenicol.
8. Dimethylsulfoxide (DMSO).

2.5. Screening Metagenomic Libraries for Cellulose Degrading Enzymatic Activity

1. CMC agar plates: 0.4% carboxymethyl cellulose (CMC) is dissolved in water by short-time heating. LB plates supplemented with 0.2% CMC are prepared by adding the 0.4% CMC solution to the dry ingredients of the medium prior to sterilization by autoclaving.
2. 48er or 96er steel stamps.
3. Congo Red solution (Roth, Karlsruhe).
4. 1 M NaCl solution.
5. 1 M HCl.

3. Methods

3.1. Sampling Procedures

3.1.1. Marine Surface Water Sampling

Surface water can be collected by either membrane pumps or any other highly effective clean pumping system on board. Further, samples can also be taken by a CTD, equipped with a 24 Niskin 10 L bottle rosette (Fig. 1). Samples from the potentially high productive surface layer around chlorophyll maxima should exceed a volume of 100 L but do not necessarily need to be larger



Fig. 1. CTD equipped with a 24 Niskin 10 L bottle rosette on German research vessel Meteor.

than 200 L, due to the high abundance of microorganisms there. After collecting, prefiltration with filters of 10 µm pore size is performed directly followed by a consecutive filtration with polycarbonate or polyvinylidenfluoride membrane filters of 0.22 µm pore size (see Note 3). To carry out this large volume filtration in an appropriate time frame, an efficient pumping system is requested, for example a peristaltic pump (see Note 2). Filters are immediately frozen and stored at –80°C (see Note 4).

3.1.2. Marine Deep Water Sampling

Samples from below the euphotic zone, where not much cell material is present, should be collected in larger volumes of at least 200 L. A CTD equipped with a 24 Niskin 10 L bottle rosette can be used for the collection of such samples; filtration is then carried out as described above. As this sampling method is limited to a certain volume, mostly 240 L, it is highly time consuming and may lead to stress responses due to dramatically changing environmental conditions during the filtration time on board (light, temperature, pressure). In this case, a sample collection by in situ pumps should be preferred. Those pumps can be set at the depth of interest, depending on the cable length of the ships' winch (Fig. 2a); this method further allows simultaneous deployment of several pumps at different depth. Therefore, the use of in situ pumps is highly time saving, and additionally leads to a higher conservation and consequently to a more realistic image of the microbial community (see Note 1). Moreover, a filtration of higher volumes of water is possible, depending on the pump type up to 5,000 L. Filtration is also conducted using carbonate membrane filters of 0.22 µm pore size, and a prefiltration is not required. After recovering the pumps, filters are immediately removed from the pumps (Fig. 2b), frozen, and stored as described above.

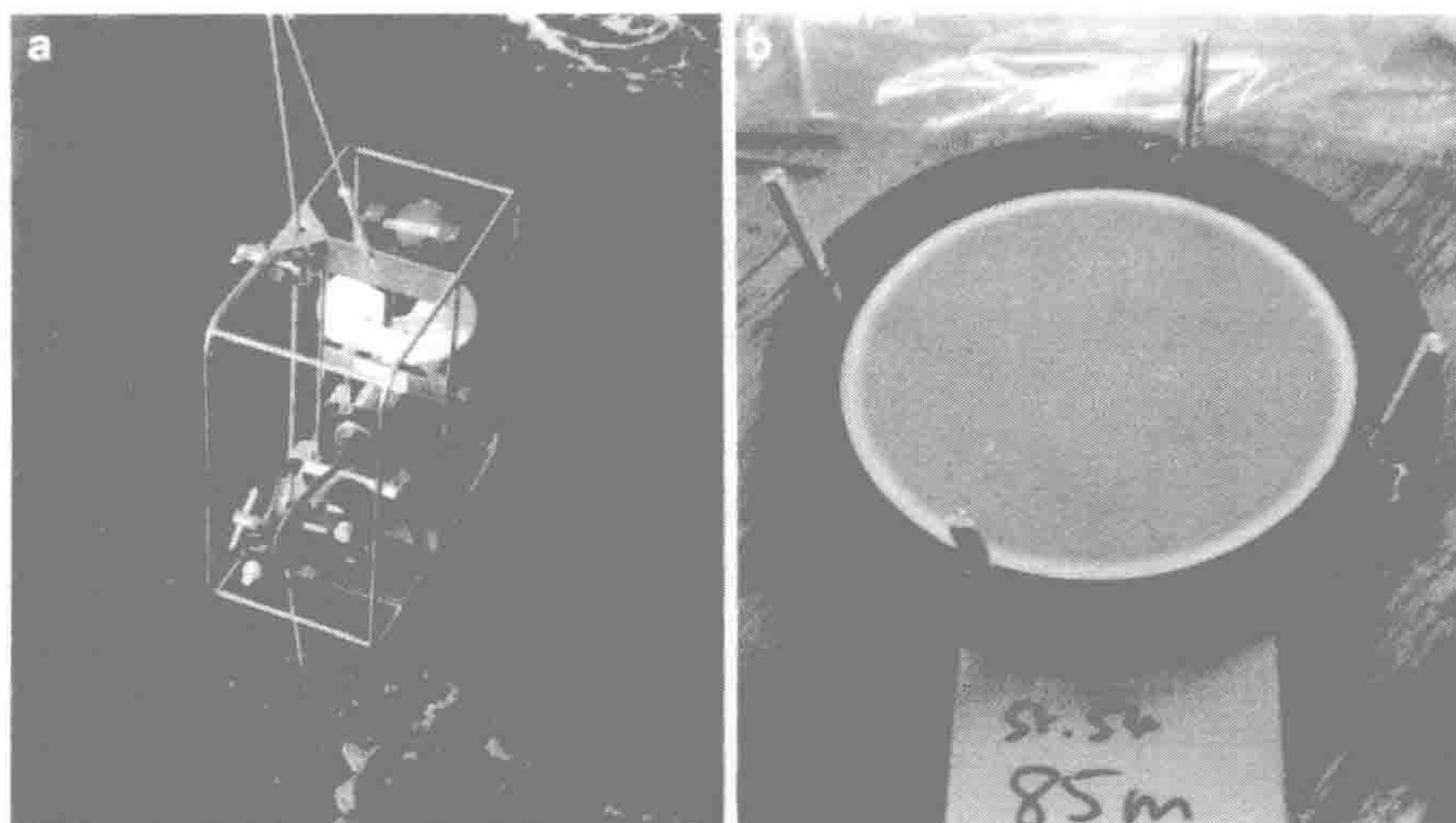


Fig. 2. Deployment of an in situ pump from RV Meteor (**a**), filter-holder with filter of an in situ pump (**b**).

3.1.3. Sampling from Marine Invertebrates

After sampling, the marine organisms are thoroughly rinsed with filtered (0.22 μm) and autoclaved seawater to remove loosely attached microorganisms. If possible the organisms are then placed in sterile Petri dishes and an area of approximately 2–5 cm^2 (depending on the amounts of microbes and downstream applications) is swabbed with a sterile cotton-tipped applicator. In the case of a fragile organism, the complete animal can be extracted for DNA isolation, resulting in a mixture of prokaryotic and eukaryotic DNA of unknown ratio. In this case, enrichment of prokaryotic cells, e.g., by fractionated centrifugation can be applied prior to DNA extraction. For comparative phylogenetic analysis, ambient seawater should be sampled and filtered as described above.

3.2. Isolation of Metagenomic DNA

DNA from filters or swabs is commonly extracted by a direct lysis of the microorganisms. Additional steps prior the lysis may be required to isolate DNA from inhibitor-contaminated habitats or enrich prokaryotic cells in order to minimize coextraction of eukaryotic DNA (17). The following modified protocol of Henne et al. (18) describes the genomic DNA isolation based on direct lysis of the microorganisms from filter or swab samples. The volumes are appropriate for 2.5 cm^2 of a filter and should be adjusted according to the filter or sample size.

1. 1.35 mL DNA extraction buffer (see Note 7), supplemented with 20 μL Proteinase K (20 mg/mL) and 200 μL lysozyme (50 mg/mL) are added to the sample followed by an incubation at 37°C for 30 min; optional shaking (150 rpm).
2. 1.5 μL (17,000 U) RNase A are added followed by further incubation at 37°C for 30 min.
3. 150 μL 20% SDS are added followed by an incubation for 2 h at 65°C and subsequent centrifugation at 4,500 × *g* for 10 min.
4. Chloroform extraction of the supernatant followed by precipitation of the nucleic acids with isopropanol (0.7 vol) for 1 h at room temperature and subsequent centrifugation for 45 min at 16,000 × *g* and 4°C.
5. The DNA precipitate is washed with 70% ethanol, dried, and solved in 25 μL TE buffer.

This extraction protocol uses enzymatic methods to remove cell walls, resulting in spheroplasts or protoplasts. The use of SDS disrupts mainly tertiary or quaternary protein structures; CTAB additionally removes polysaccharides and remaining proteins. An increase from 1 to 5% CTAB in the DNA extraction buffer allows an improved lysis of archaeal cell walls, which significantly differ from the bacterial cell walls (19, 20) (see Note 5). In some cases, e.g., DNA extraction of samples containing high amounts of

Gram-positive bacteria, initial mechanical cell lyses might be necessary, e.g., using a bead beater with small glass, ceramic, zirconium, or steel beads (21) (see Note 6). Finally, the isolated metagenomic DNA is analyzed by gel electrophoresis and should contain large fragments (Fig. 3) in case of constructing a metagenomic large-insert library.

3.3. 16S rDNA Phylogenetic Analysis

Bacterial and archaeal 16S rRNA genes present in the metagenomic DNA are commonly PCR amplified using the bacteria-specific primer 27F and the universal primer 1492R (22) or the archaea-specific primer set 20F and 958R (23), 2–10 ng of extracted DNA (see Note 8) and a standard amplification protocol; e.g., 5 min at 95°C followed by 30 cycles of 30 s at 94°C, 45 s at 55°C, and 1.5 min at 72°C. The resulting 1,500 bp bacterial or 1,000 bp archaeal PCR fragments (Fig. 4) are purified and

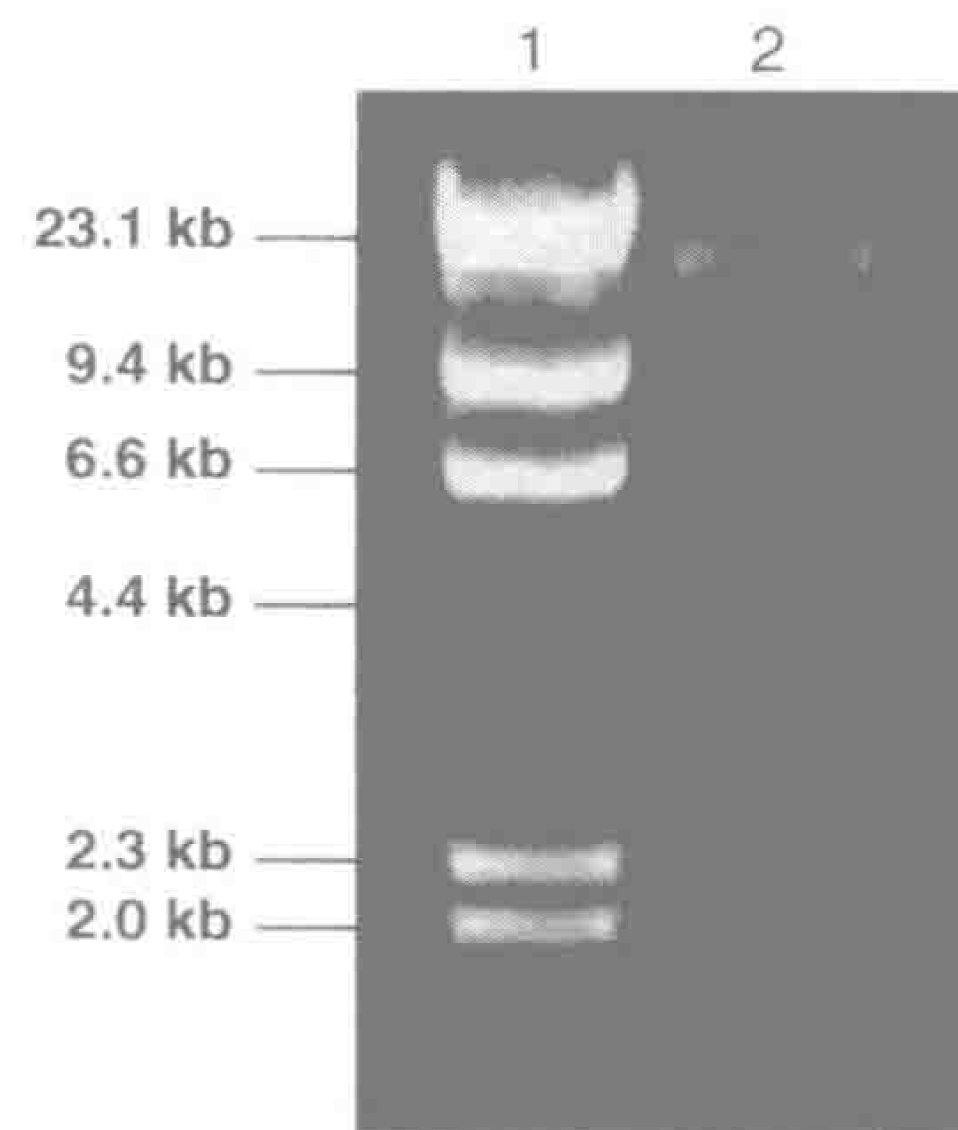


Fig. 3. Gel electrophoretic analysis of metagenomic high molecular weight DNA.

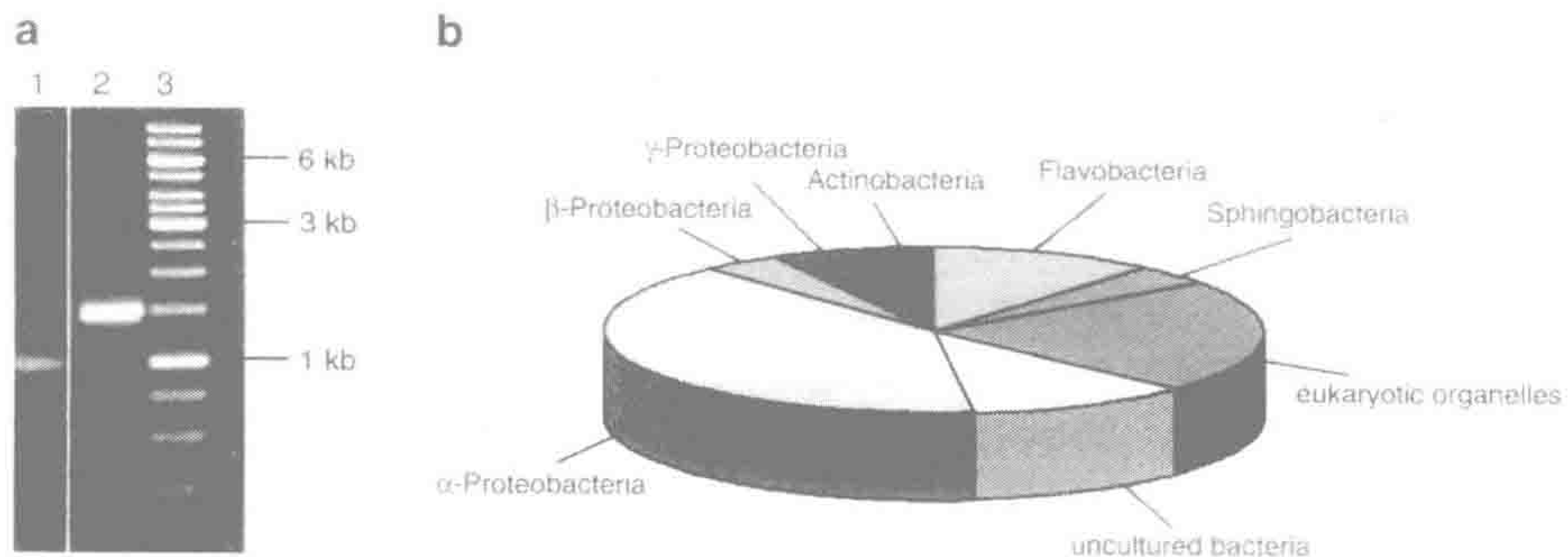


Fig. 4. 16S rDNA phylogenetic analysis of a marine habitat. (**a**) 16S rDNA gene amplification of metagenomic DNA using universal bacterial (*line 2*) and archaeal (*line 1*) primers. (**b**) Respective phylogenetic composition of the marine habitat based on 16S rDNA sequencing analysis.

cloned into a sequencing vector, e.g., a TA cloning vector allowing an efficient cloning (24), followed by independent and complete DNA sequence analyses for both strands using the primers 27F and 1492R or 20F and 958R or universal primers complementary to the flanking vector regions (25, 26). The 16S rDNA analysis not only allows insight into present community structure of the respective habitat, but also points out the likely potential of the habitat to detect new biotechnological relevant enzymes. In addition to the knowledge gained on the actual microbial diversity, additional PCR amplifications can be performed using specific primer sets in order to analyze the presence of functional genes, e.g., the *nifH* gene for diazotrophes, encoding a structural gene of nitrogenase, the key enzyme of nitrogen fixation (27, 28).

3.4. Construction of a Metagenomic Large-Insert Library

Fosmid and bacterial artificial chromosome (BAC) vectors have been developed to clone large genomic DNA fragments of up to 40 kb and ~120 kb, respectively. These vectors replicate using the single-copy F-factor replicon and show high stability carrying large inserts (29). Recently, novel large-insert vectors have been developed carrying both the single-copy and an additional inducible high copy number origin of replication (11). This ensures on the one hand insert stability and successful cloning of encoded and expressed toxic proteins and unstable DNA sequences, and on the other hand allows increased DNA yields in vector preparations and functional screens of clone libraries by induction to high copy numbers (30). Thus, BACs and fosmids have become standard tools for constructing genomic clone libraries. Genomic library construction kits are commercially available that pursue blunt-end cloning strategies resulting in complete and unbiased libraries. The "Copy Control™ Fosmid Library Production Kit" (e.g., with pCC1FOS) combines all advantages to stable insert large DNA fragments into the vector with little expenditure of time (Fig. 5). In the following, the corresponding protocol according to manufacturer's instructions is presented.

1. *Preparation of DNA*: High molecular weight (meta)genomic DNA is isolated as described above and diluted in TE buffer at a concentration of ~500 ng/μL (see Note 9).
2. *Shearing*: DNA fragments in the range of 20–40 kb are obtained by multiple pipetting the DNA solution using a 1,000 μL pipette tip.
3. *End-repair of the DNA fragments*: The end-repair reaction described below generates blunt-ended, 5′-phosphorylated DNA fragments and can be scaled up or down depending on the amount of available DNA, followed by incubation at room temperature (RT) for 45 min (see Note 10).

Sterile water	× μL
10× End-repair buffer	8 μL
2.5 mM dNTPs	8 μL
10 mM ATP	8 μL
Up to 20 μg sheared DNA	× μL
End-repair enzyme mix	4 μL
Total reaction volume	80 μL

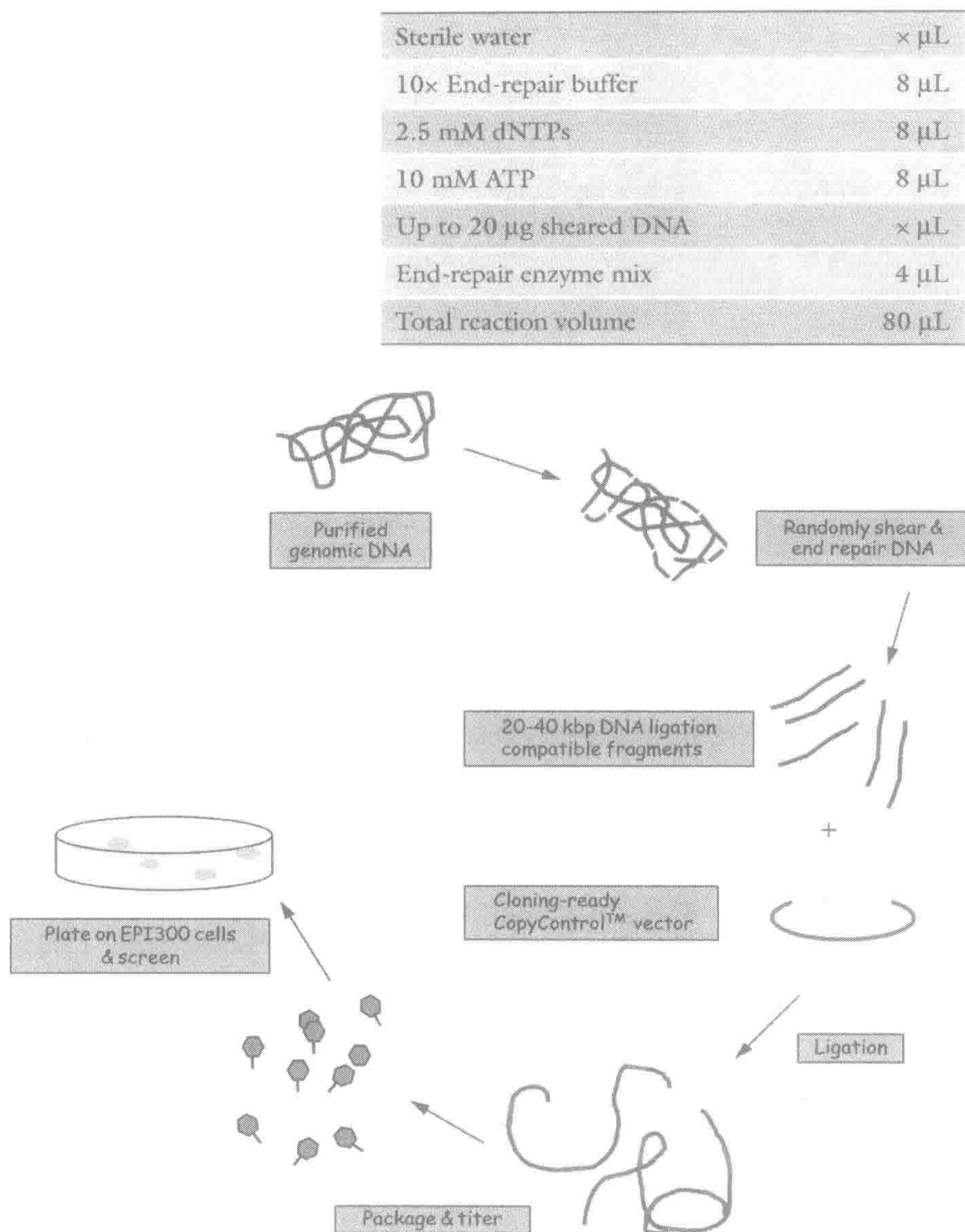


Fig. 5. Construction of a metagenomic library (modified according to Epicentre, Madison/USA).

4. *Dialysis*: The end-repair reaction mix is dialyzed for 30 min at RT against sterile water to remove interfering salts. This step can be performed, e.g., by using 0.025 μm cellulose filters type VS placed on the surface of sterile water in a Petri dish, on which the reaction mix is placed.

5. *Ligation*: The ligation reaction is mixed in a 10:1 molar ratio of *CopyControl pCC1FOS vector* to insert DNA and incubated for 2 h at RT followed by overnight incubation at 16°C (see Note 11). The following reagents are combined in the order listed.

Sterile water	× μL
10× Fast-link ligation buffer	1 μL
10 mM ATP	1 μL
CopyControl pCC1FOS vector (0.5 mg/mL)	1 μL
Insert DNA (0.25 μg of 40 kb DNA)	× μL
Fast-link DNA ligase	1 μL
Total reaction volume	10 μL

6. *Packaging reaction*: 10 μL of the ligation reaction are added to one-half of the provided *MaxPlax Lambda Packaging extract* (25 μL) in a reaction tube being kept on ice. The packaging reaction is incubated at 30°C. After 90 min, the remaining 25 μL of *Lambda Packaging Extract* are added and the reaction is incubated for additional 90 min at 30°C. Following the incubation, the *Phage-Dilution buffer* is added to 1 mL final volume and mixed gently. For storage at 4°C, 25 μL of chloroform are added.
7. *Titration of the packaged CopyControl fosmid library*: Prior transducing the complete packaging reaction, it is recommended to determine the phage particle titer (e.g., CopyControl Fosmid clones). Ten microliters of the packaging reaction is added to 100 μL of exponentially growing EPI300-T1^R host cells (LB containing 10 mM $MgSO_4$) followed by incubation at 37°C for 20 min. Aliquots of the transduced EPI300-T1^R cells are plated on LB plates supplemented with 12.5 μg/mL chloramphenicol and incubated overnight at 37°C to select for the CopyControl Fosmid clones. Colonies are counted and the phage particles titer is calculated.
8. *Transduction and plating the CopyControl fosmid library*: According to the titration and the estimated number of clones required, the volume of the packing reaction (fosmid library) required for the construction of the respective clone library is calculated. The transduction into EPI300-T1^R host cells is performed as described above in several parallel reactions using the volumes mentioned above. Appropriate aliquots of the infected bacteria are plated on LB plates supplemented with 12.5 μg/mL chloramphenicol for selection and incubated overnight at 37°C. Fosmid clones obtained are grown in microtiter plates (96 wells) and subsequently stored at –70°C in the presence of 8% DMSO.
9. *Induction to higher copy numbers*: The fosmid clones of a library can be induced to reach higher fosmid copy numbers in order to achieve high fosmid DNA yields for sequencing, fingerprinting, or other downstream applications. Induction to

higher copy numbers is also recommended for direct function-based screening assays of the clone library, e.g., on plates. The induction can be achieved in any desired culture volume depending on the downstream application. In general, LB medium is supplemented with chloramphenicol and 1:1,000 of induction solution and 1:10 of the respective overnight culture followed by incubation for 5 h at 37°C with agitation.

3.5. Sequence-Based Screens of Metagenomic Libraries Using a PCR-Amplification Approach

A sequence-based analysis of metagenomic DNA can be performed by monitoring the presence of respective key genes by PCR amplification to identify genes and metabolic pathways. The primers are designed based on the sequences known for the respective gene with the primers binding to conserved regions of the genes. PCR amplification is performed using the metagenomic DNA, fosmid pools, or single fosmids of the metagenomic library. The respective amplified PCR fragment is cloned (e.g., into a TA cloning vector) followed by sequence analysis of randomly chosen clones. A recent example is the identification of a gene encoding a novel cytochrome P450 monooxygenase with a robust catalytic activity in a soil metagenomic library (31). Another example is the unexpected high diversity and distribution of the *nifH* gene, one of the functional key genes for nitrogen fixation, discovered in the surface water of the Pacific Ocean (27, 28). Large-scale sequencing projects such as the one initiated by Craig Venter for the metagenome of the Saragossa Sea resulted in the identification of numerous novel genes and is a famous example of sequence-based metagenome analyses (32). Recently, large-scale sequencing of complete metagenomes by massive parallel sequencing, e.g., a pyro-sequencing approach, has been performed followed by bioinformatic analyses and partial assembly of the genomes present in the habitats (33, 34).

3.6. Function-Based Screens of Metagenomic Libraries

Functional screens for novel genes in metagenomic libraries explore the genetic potential of a habitat by directly monitoring products or enzymatic activity of the metagenomic clones. Metagenomic libraries have been screened for various biomolecules, such as biotechnologically relevant enzymes. So far, functional screens of metagenomic libraries have identified, e.g., several novel antibiotics (e.g., turbomycin A and B) (35), aminoacylated antibiotics (36), or small antimicrobial molecules (37) from soil metagenomes, exoenzymes such as lipases (38) and marine chitinases (39) or membrane proteins (40). In the following, the screen for cellulose degrading activity will be exemplarily described.

3.6.1. Screening Metagenomic Libraries for Novel Cellulose Degrading Enzymatic Activity

Cellulases refer to a class of enzymes that catalyze the hydrolysis of cellulose by hydrolyzing the 1,4-β-D-glycosidic linkages. Several different kinds of cellulases are known that differ structurally and mechanistically (41). Cellulases are widely used in the pulp

and paper industry for various purposes, for pharmaceutical applications (42), in the textile industry, and are present in laundry detergents. Thus, these enzymes have a high relevance in biotechnology and novel thermal or pH-stable cellulases are continuously searched for industrial application. Metagenomic libraries can be screened for cellulose degrading activities on agar plates with a rapid and sensitive assay system using Congo Red for detection. As Congo Red shows a strong interaction with polysaccharides containing contiguous β-(1,4)-linked D-glucopyranosyl units and a significant interaction with β-(1,3)-D-glucans, zones of cellulose hydrolysis around a metagenomic clone on agar plates can be visualized (43). The following protocol exemplarily describes a plate screen for cellulose degrading enzymatic activity on plates.

1. Metagenomic clones stored at –70°C are directly transferred from the 96-well microtiter plates to the CMC agar plates with a steel stamp, followed by incubation for 24 h at 37°C.
2. The plates are flooded with an aqueous 0.2% solution of Congo Red for 15 min.
3. After pouring off the Congo Red solution, the plates are further treated by flooding with 1 M NaCl for 15 min. Degradation of cellulose is indicated by a yellow zone around positive clones (see Fig. 6).
4. The visualized zones of hydrolysis can be stabilized for at least 2 weeks by additional flooding the agar plate with 1 M HCl, which changes the dye color to blue and inhibits further enzyme activity.

In order to identify the respective open reading frame (ORF) of a confirmed fosmid conferring the desired activity, an in vitro transposon mutagenesis can be performed, e.g., using the EZ-Tn5™<oriV/KAN-2> Insertion Kit from Epicentre (Madison/USA). Following the transposon mutagenesis clones

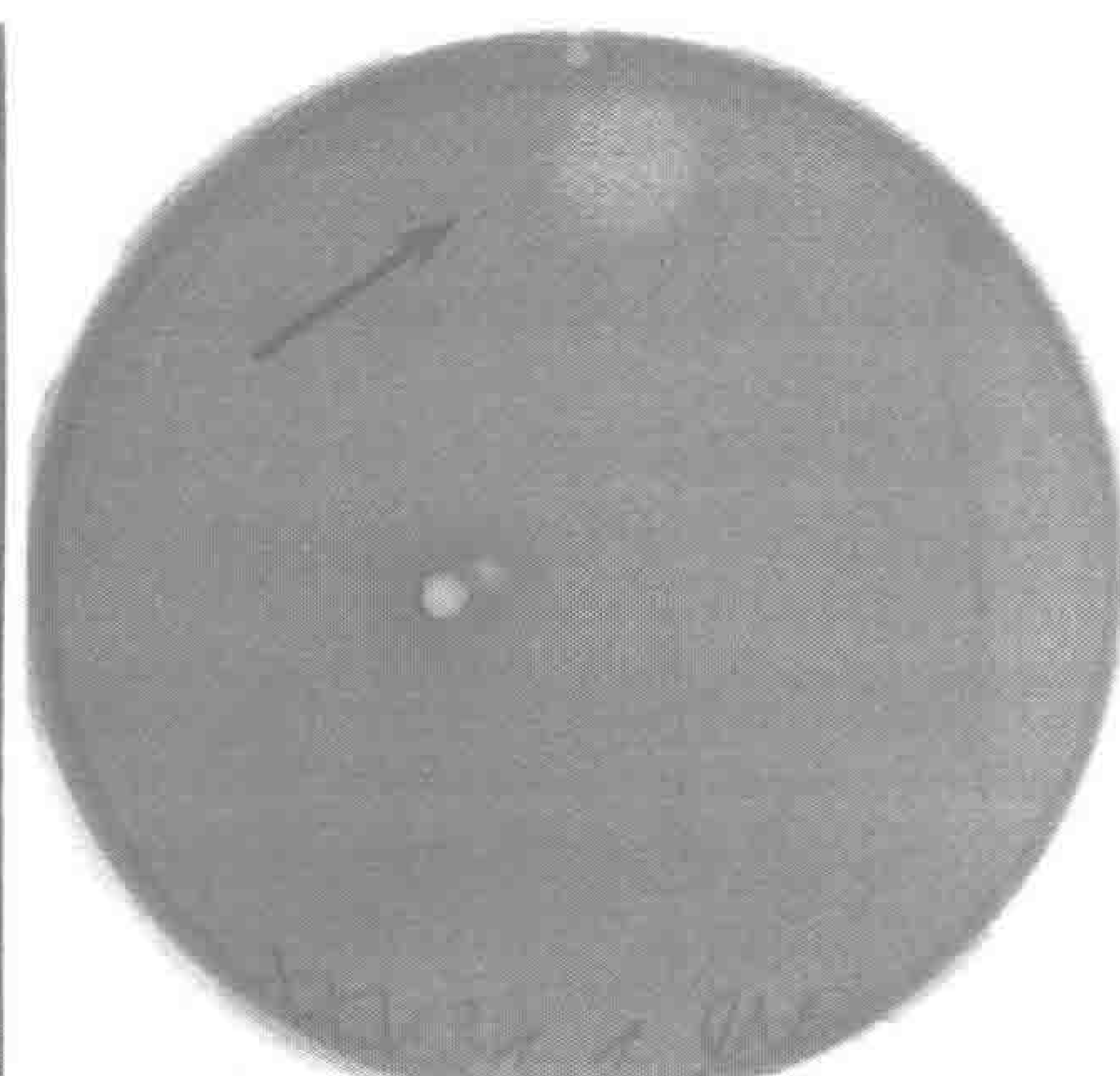


Fig. 6. Plate screen for cellulose degrading metagenomic clones.

are screened for loss of the desired activity. Fosmid DNA of clones that lost the activity are sequenced using primers hybridizing to the 5′ and 3′ end of the transposon reading into the flanking metagenomic regions. The obtained DNA sequences flanking the transposon are assembled in order to identify the respective ORF, which can be cloned in an expression vector to purify the protein in high amounts.

4. Notes

1. Based on the higher number of pores polyvinylidenfluoride filters are preferred to filter high water volumes through a single filter especially when working with small filter diameters.
2. The filtration should be realized as fast as possible with a supporting peristaltic pump in a cold room or for large volumes preferentially using an in situ pump at the respective conditions at the sampling site.
3. For long-time storage, filters have to be frozen at −80°C in practicable dimensions. Before liquid nitrogen treatment, the filters have to be cut into convenient pieces to rule out needless freeze/thaw cycles.
4. The sampling procedures have to be performed rapidly because of the changing environmental conditions.
5. The standard DNA extraction protocol has to be modified when the samples contain high amounts of polysaccharides and glycoproteins. In this case, the sample should be treated with higher percentages of CTAB to support disintegration of samples.
6. Sometimes extracted metagenomic DNA shows a high degradation because of DNases present in the sample. In this case, the addition of EDTA to the DNA extraction buffer helps suppressing the damage of DNA. (EDTA is used for scavenging metal ions to deactivate metal-dependent enzymes).
7. In some cases, an additional mechanical cell lyses step might be necessary as some bacteria/archaea may not be cracked with enzymatic methods.
8. The crucial step of the 16S rDNA PCR amplification is to amplify the bacterial/archaeal 16S rDNA fragments from the optimal amount of template DNA, which can differ from 1 pg to 1 μg.
9. If the extracted metagenomic DNA will be used for library construction, the DNA should routinely be analyzed for degradation to decide if shearing is necessary or this step might be skipped.

10. Before preparing the end-repair reaction, the DNA concentration has to be determined precisely by measuring the absorbance at 260/280 nm, as in the following dNTPs are added and all following steps and calculations are based on this DNA quantification.
11. A size selection of 20–40 kb end-repaired fragments can be performed to ensure that only large inserts are ligated into the pCC1FOS vector. In special cases, the molar ratio 10:1 of fosmid vector to insert DNA can be optimized (5:1 or 7.5:1) to increase the clone number.

References

1. Amann, R.I., Ludwig, W., and Schleifer, K.-H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59, 143–169.
2. Streit, W.R. and Schmitz, R.A. (2004) Metagenomics – the key to the uncultured microbes. *Curr Opin Microbiol* 7, 492–498.
3. Lorenz, P. and Eck, J. (2005) Metagenomics and industrial applications. *Nat Rev Microbiol* 3, 510–516.
4. Pham, V.D., Palden, T., and DeLong, E.F. (2007) Large-scale screens of metagenomic libraries. *J Vis Exp* (4), 201.
5. DeLong, E.F. (2009) The microbial ocean from genomes to biomes. *Nature* 459, 200–206.
6. Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 68, 669–685.
7. Beja, O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A., Nguyen, L.P., et al. (2000) Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol* 2, 516–529.
8. Beja, O., Spudich, E., Spudich, J., Leclerc, M., and DeLong, E. (2001) Proteorhodopsin phototrophy in the ocean. *Nature* 411, 786–789.
9. de la Torre, J.R., Christianson, L.M., Beja, O., Suzuki, M.T., Karl, D.M., Heidelberg, J., et al. (2003) Proteorhodopsin genes are distributed among divergent marine bacterial taxa. *Proc Natl Acad Sci USA* 100, 12830–12835.
10. Woyke, T., Teeling, H., Ivanova, N.N., Huntemann, M., Richter, M., Gloeckner, F.O., et al. (2006) Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* 443, 950–955.
11. Wild, J., Hradecna, Z., and Szybalski, W. (2002) Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. *Genome Res* 12, 1434–1444.
12. Shizuya, H. and Kouros-Mehr, H. (2001) The development and applications of the bacterial artificial chromosome cloning system. *Keio J Med* 50, 26–30.
13. Azam, F. (1998) Microbial control of oceanic carbon flux: the plot thickens. *Science* 280, 694–696.
14. DeLong, E.F. and Karl, D.M. (2005) Genomic perspectives in microbial oceanography. *Nature* 437, 336–342.
15. Karl, D.M. (2007) Microbial oceanography: paradigms, processes and promise. *Nat Rev Microbiol* 5, 759–769.
16. Kennedy, J., Marchesi, J.R., and Dobson, A.D. (2007) Metagenomic approaches to exploit the biotechnological potential of the microbial consortia of marine sponges. *Appl Microbiol Biotechnol* 75, 11–20.
17. Gabor, E.M., de Vries, E.J., and Janssen, D.B. (2003) Efficient recovery of environmental DNA for expression cloning by indirect extraction methods. *FEMS Microbiol Ecol* 44, 153–163.
18. Henne, A., Daniel, R., Schmitz, R.A., and Gottschalk, G. (1999) Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl Environ Microbiol* 65, 3901–3907.
19. Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., et al. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci USA* 103, 12115–12120.
20. De Corte, D., Yokokawa, T., Varela, M.M., Agogue, H., and Herndl, G.J. (2009) Spatial distribution of Bacteria and Archaea and *amoA* gene copy numbers throughout the water column of the Eastern Mediterranean Sea. *ISME J* 3, 147–158.

21. Treusch, A.H., Kletzin, A., Raddatz, G., Ochsenreiter, T., Quaiser, A., Meurer, G., et al. (2004) Characterization of large-insert DNA libraries from soil for environmental genomic studies of Archaea. *Environ Microbiol* 6, 970–980.
22. Lane, B.G., Bernier, F., Dratewka-Kos, E., Shafai, R., Kennedy, T.D., Pyne, C., et al. (1991) Homologies between members of the germin gene family in hexaploid wheat and similarities between these wheat germins and certain Physarum spherulins. *J Biol Chem* 266, 10461–10469.
23. DeLong, E.F. (1992) Archaea in coastal marine environments. *Proc Natl Acad Sci USA* 89, 5685–5689.
24. Mead, D.A., Pey, N.K., Herrnstadt, C., Marcil, R.A., and Smith, L.M. (1991) A universal method for the direct cloning of PCR amplified nucleic acid. *Biotechnology (N Y)* 9, 657–663.
25. Pages, R.D.M. and Holmes, E.C. (1998) Molecular Evolution: A Phylogenetic Approach. Blackwell, Oxford.
26. Suarez-Diaz, E. and Anaya-Munoz, V.H. (2008) History, objectivity, and the construction of molecular phylogenies. *Stud Hist Philos Biol Biomed Sci* 39, 451–468.
27. Langlois, R.J., LaRoche, J., and Raab, P.A. (2005) Diazotrophic diversity and distribution in the tropical and subtropical Atlantic Ocean. *Appl Environ Microbiol* 71, 7910–7919.
28. Langlois, R.J., Hummer, D., and LaRoche, J. (2008) Abundances and distributions of the dominant *nifH* phylotypes in the Northern Atlantic Ocean. *Appl Environ Microbiol* 74, 1922–1931.
29. Wild, J., Hradecna, Z., Posfai, G., and Szybalski, W. (1996) A broad-host-range *in vivo* pop-out and amplification system for generating large quantities of 50- to 100-kb genomic fragments for direct DNA sequencing. *Gene* 179, 181–188.
30. Sektas, M. and Szybalski, W. (1998) Tightly controlled two-stage expression vectors employing the Flp/FRT-mediated inversion of cloned genes. *Mol Biotechnol* 9, 17–24.
31. Kim, B.S., Kim, S.Y., Park, J., Park, W., Hwang, K.Y., Yoon, Y.J., et al. (2007) Sequence-based screening for self-sufficient P450 monooxygenase from a metagenome library. *J Appl Microbiol* 102, 1392–1400.
32. Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., et al. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66–74.
33. Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., et al. (2005) Comparative metagenomics of microbial communities. *Science* 308, 554–557.
34. Green, B.D. and Keller, M. (2006) Capturing the uncultivated majority. *Curr Opin Biotechnol* 17, 236–240.
35. Gillespie, D.E., Brady, S.F., Bettermann, A.D., Cianciotto, N.P., Liles, M.R., Rondon, M.R., et al. (2002) Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Appl Environ Microbiol* 68, 4301–4306.
36. Brady, S.F., Chao, C.J., and Clardy, J. (2002) New natural product families from an environmental DNA (eDNA) gene cluster. *J Am Chem Soc* 124, 9968–9969.
37. MacNeil, I.A., Tiong, C.L., Minor, C., August, P.R., Grossman, T.H., Loiacono, K.A., et al. (2001) Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J Mol Microbiol Biotechnol* 3, 301–308.
38. Henne, A., Schmitz, R.A., Bomeke, M., Gottschalk, G., and Daniel, R. (2000) Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl Environ Microbiol* 66, 3113–3116.
39. Cottrell, M.T., Moore, J.A., and Kirchman, D.L. (1999) Chitinases from uncultured marine microorganisms. *Appl Environ Microbiol* 65, 2553–2557.
40. Majernik, A., Gottschalk, G., and Daniel, R. (2001) Screening of environmental DNA libraries for the presence of genes conferring Na(+)(Li(+))/H(+) antiporter activity on *Escherichia coli*: characterization of the recovered genes and the corresponding gene products. *J Bacteriol* 183, 6645–6653.
41. Ghose, T.K. (1987) Measurement of cellulase activities. *Pure Appl Chem* 59, 257–268.
42. Cohen, S. (2007) Bezoars and foreign bodies. In: Porter, R.S., Kaplan, J.L., Homeier, B.P., and Beers, M.H. Eds. The Merck Manual-Online Library. http://www.merck.com/mmpe/sec02/ch014/ch014c.html
43. Teather, R.M. and Wood, P.J. (1982) Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Appl Environ Microbiol* 43, 777–780.

Chapter 4

Metagenomic Analysis of Isotopically Enriched DNA

Yin Chen, Josh D. Neufeld, Marc G. Dumont, Michael W. Friedrich, and J. Colin Murrell

Abstract

This detailed protocol describes an approach for combining DNA stable-isotope probing-based enrichment, multiple displacement amplification (MDA), and metagenomics. Together, these three methodologies enable selective access to the genomes of uncultivated organisms that actively grow using isotopically labelled carbon and nitrogen sources. Incubations with stable-isotope-labelled substrates enrich isotopically labelled DNA from functionally relevant micro-organisms; this serves as a filter to reduce the complexity of the metagenome. The MDA step generates sufficient DNA from labelled nucleic acid for metagenomic library construction. Subsequently, genome fragments can be subjected to a variety of screens for phylogenetic or functional genes relevant to active community members. The MDA-generated DNA can also serve as template for direct high-throughput sequencing to aid reconstruction of metabolic pathways of those active organisms. Recent proof-of-concept studies have demonstrated that this novel combination of molecular methods can offer substantial enhancements to gene detection frequencies and may have great future potential for the discovery of novel genes, enzymes, and metabolic pathways.

Key words: DNA stable-isotope probing, Multiple displacement amplification, Metagenomics

1. Introduction

Molecular ecology has revolutionised our understanding of microbial diversity in the environment over the last two decades as this approach has bypassed the limitations inherent to classical cultivation strategies. More than 600,000 sequences of the ubiquitous marker SSU rRNA have been collected to date by cultivation-independent surveys. And yet, the physiology and functions of these uncultivated micro-organisms represented only by 16S rRNA gene sequences are largely unknown. One successful strategy to unravel the function of uncultivated

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668, DOI 10.1007/978-1-60761-823-2_4, © Springer Science+Business Media, LLC 2010

microbes is metagenomics, which represents the retrieval and analysis of genome fragments from all community members in an environmental sample (1). Metagenomic DNA containing the 16S rRNA gene can reveal the phylogenetic affiliation of the original micro-organism as well as adjacent enzyme-encoding genes. These "functional" genes help deduce the potential role of these micro-organisms in the environment. Conventional metagenomic approaches involve cloning of extracted environmental DNA followed by sequence-based and/or function-based screens. However, given the relative rarity of most microbial community members, massive shotgun sequencing of environmental DNA focuses on the most abundant species in a given sample. For example, given an assumed underlying community structure for the Global Ocean Sampling expedition dataset (2), only ~50% of the total community DNA has been captured, despite substantial sequencing effort; nearly five times the sequencing reads would be required to access 90% of the diversity of these samples (3). Furthermore, function-based screening for enzymes of relevance to industry, biotechnology, and pharmaceutics may be challenging due to extremely low-target gene frequency. An alternative approach might be classical enrichment of individual populations, and thereby the enrichment of genes of interest prior to metagenomic analysis. Enrichment often results in the selection of micro-organisms that are irrelevant to the natural environment but adapted best to the enrichment conditions.

Stable-isotope probing of DNA (DNA-SIP) is a cultivation-independent method to selectively label micro-organisms that can metabolise a specific stable-isotope-labelled substrate [*e.g.* ^{13}C, ^{15}N; (4)]. Since its development, DNA-SIP has been widely used to study micro-organisms involved in particular bioprocesses [reviewed in (5) and (6)]. When carried out under near-natural conditions, SIP has the potential to select for labelled genomes of active populations while minimising the extent of enrichment bias (7). In combination with metagenomics, DNA-SIP facilitates the selective isolation of DNA from functionally relevant micro-organisms to construct metagenomic libraries in a directed fashion that has not been possible previously (5, 7–10). This goal was first achieved by exposing a soil sample to $^{13}CH_4$, retrieving high-quality labelled DNA without damaging UV exposure, and generating a modest BAC library for the discovery of multiple clones with *pmoA*-containing operons (11). However, the major concern of combining DNA-SIP with metagenomics has been the challenge of obtaining sufficient ^{13}C-labelled "heavy" DNA for construction of a metagenomic library, without using excessively high concentrations of substrate and reducing the risk of enrichment bias. This has been recently overcome by applying multiple displacement amplification [MDA; (12–14)] to gener-

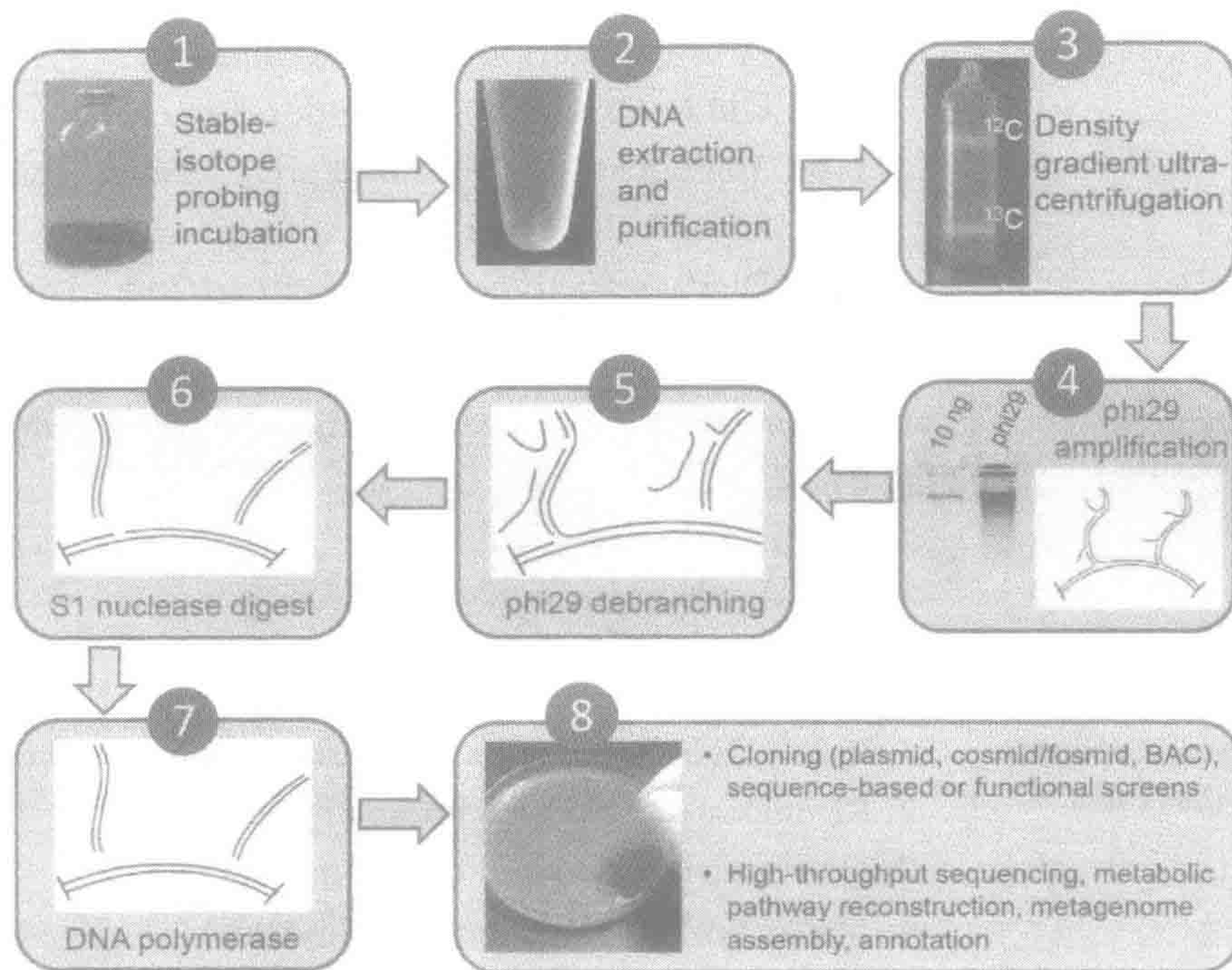


Fig. 1. Overview of the combination of DNA-SIP, multiple displacement amplification, and metagenomics. Diagrams for enzymatic treatment of MDA products were modified with permission from (23).

ate large quantities of DNA from minute quantities of ^{13}C-labelled "heavy" DNA (13, 14). Outlining this novel approach, we present here a detailed protocol for using MDA to prepare metagenomic libraries from DNA that has been "filtered" by SIP pre-incubation (Fig. 1).

2. Materials

2.1. Reagents and Equipments for DNA-SIP

The reagents and equipment required for DNA-SIP have been described elsewhere in detail, and we recommend readers to refer to these protocols (15, 16). Briefly, reagents needed include:

1. Suitable stable-isotope (i.e. ^{13}C or ^{15}N) -labelled compounds.
2. CsCl stock solution (density 1.890 g/mL in water).
3. Gradient buffer: 0.1 M Tris–HCl, 0.1 M KCl, 1 mM EDTA, pH 8.0.
4. DNA precipitation buffer: 30% polyethylene glycol 6000, 1.6 M NaCl.
5. Glycogen: 20 mg/mL in water.
6. TE buffer: 10 mM Tris–HCl, 1 mM EDTA, pH 7.6.

7. The key instruments for DNA-SIP include:
   (a) Ultracentrifuge and corresponding rotor.
   (b) Appropriate device for measuring substrate consumption.
   (c) Suitable device for gradient fractionation.
   (d) Digital refractometer (*e.g.* Reichert 2000™) or an analytic balance for measuring gradient density.

2.2. Multiple Displacement Amplification

1. GenomiPhi V2 DNA Amplification Kit (GE Healthcare).
2. Thermal cycler.
3. 500 μL PCR tubes.
4. Pipettors and corresponding tips.
5. Agarose.

2.3. Enzyme Treatment of MDA-Generated DNA

1. Phi29 DNA polymerase and corresponding buffer.
2. S1 nuclease and corresponding buffer.
3. DNA polymerase I and corresponding buffer.
4. Microcon YM-30 column (Millipore).
5. 0.5 M Ethylene diamine tetraacetic acid (EDTA).
6. Phenol:chloroform:isoamyl alcohol (25:24:1 v/v; pH 8.0).
7. Chloroform:isoamyl alcohol (24:1 v/v).
8. Reagent-grade ethanol.
9. Microcentrifuge.

2.4. Metagenome Library Construction and Screening

1. CopyControl™ Fosmid Library Production Kit (Epicentre), including End-Repair enzyme mix (Epicentre) and GELase (Epicentre).
2. CHEF Mapper pulsed-field gel electrophoresis system (*e.g.* Bio-Rad).
3. Low-melting point agarose.

3. Methods

3.1. DNA Extraction and Preparation of ^{13}C-Labelled DNA

A proper DNA-SIP experiment includes the following key steps:

1. SIP incubation with stable-isotope compounds.
2. DNA extraction from the labelled environmental samples.
3. Ultracentrifugation and gradient fractionation.
4. DNA precipitation and identification of labelled "heavy" DNA.

As mentioned in the previous section, it is not our intention to present a detailed protocol for DNA-SIP set up in this chapter and we refer readers to (15) and (16) to determine appropriate incubation conditions and steps for purifying labelled DNA from caesium chloride (CsCl) gradients. When ^{13}C-labelled "heavy" DNA is purified from CsCl gradient fractions, it serves as template for subsequent amplification and cloning within metagenomic libraries (see Note 1).

3.2. Multiple Displacement Amplification

In the initial study by Dumont et al. (11), a ^{13}C-substrate concentration which exceeded that which was present *in situ* was chosen to obtain the microgram quantities of DNA required for preparing metagenomic libraries. The potential drawbacks of using elevated substrate concentrations and long incubation times include an enrichment bias and cross-feeding of substrate label to non-target populations. Here, we describe an alternative protocol that uses DNA retrieved from SIP incubations with near *in situ* concentrations of labelled substrate and short incubation times. The low yields of labelled DNA from the DNA-SIP are then augmented using MDA (Fig. 1).

1. 1 μL of "heavy" DNA (^{13}C-labelled), ideally ~1–10 ng, is mixed with 9 μL of sample buffer (see Note 2).
2. The mixture is heated to 95°C for 3 min and then placed on ice.
3. 9 μL of reaction buffer and 1 μL of phi29 enzyme mix are combined, and then added to the template and buffer mixture, which is incubated at 30°C for 1.5 h for amplification.
4. The enzyme is inactivated by heating at 65°C for 10 min.
5. 1 μL of the amplification product is loaded on a 1% (w/v) agarose gel to quantify yield (see Note 3).

3.3. Enzyme Treatment of MDA-Generated DNA

MDA-generated DNA is hyperbranched and contains both double-stranded chimeras (which are stable) and chimeras that contain single-stranded regions (which may be cleaved enzymatically); the processes leading to chimera formation have been described in (17). This study indicated that chimeras were formed by rearrangement of two neighbouring fragments (usually <10 kb in distance in the genomic sequence), of which rearrangement of inverted sequences with intervening deletions was the major chimera type during MDA (~85%). In order to resolve the hyperbranched structure and eliminate single-stranded chimeras, the reaction products may be treated enzymatically to generate double-stranded DNA suitable for cloning (Fig. 1). The enzyme treatment process outlined below was initially introduced by Zhang et al. (18) and is the most effective method tested for reducing the occurrence of chimeras (see Note 4).

1. MDA-generated DNA is purified using a Microcon YM-30 column (Millipore) and washed with water to remove random hexamer primers according to the manufacturer's directions (see Note 5).
2. The purified DNA (200 μL) is then incubated with 10 U of phi29 DNA polymerase (Fermentas) at 30°C for 2 h without adding random hexamer primers for debranching.
3. The reaction is stopped by heating at 65°C for 10 min.
4. The DNA is purified using another Microcon YM-30 column (Millipore).
5. The purified DNA is eluted by 100 μL of water and then incubated with 200 U of S1 nuclease (Fermentas) at 37°C for 30 min to digest single-stranded DNA.
6. The reaction is stopped by adding 50 μL of 0.5 M EDTA and heating at 70°C for 10 min.
7. DNA is extracted with phenol:chloroform:isoamyl alcohol (25:24:1, v/v), again with chloroform:isoamyl alcohol (24:1, v/v), and then precipitated with ethanol (see Note 6).
8. The pellet is air-dried and dissolved in 100 μL of nuclease-free water, to which 20 U of DNA polymerase I and 0.4 μL dNTP mix (25 mM each) are added for nick translation to repair gaps within the DNA.
9. The reaction is incubated at 25°C for 1 h and then stopped by heating at 75°C for 10 min.

3.4. Metagenome Library Construction and Screening

The enzyme-treated, MDA-generated DNA can be used for high-throughput sequencing (*e.g.* 454 pyrosequencing) or for construction of a metagenomic clone library (*e.g.* plasmid, cosmid/fosmid, or BAC vectors). Readers are recommended to refer to Mußmann et al. (19) for protocol details involving DNA preparation for direct high-throughput sequencing, which is not the focus of this book chapter. Here, we describe a protocol for making a fosmid library using the enzyme-treated, MDA-generated DNA.

1. After DNA polymerase I treatment and heat inactivation of the enzyme, the DNA is end-repaired to generate blunt ends using the End-Repair enzyme mix.
2. The end-repaired DNA is then loaded onto a 1% (w/v) low-melting point agarose gel for size selection using a CHEF Mapper pulsed-field gel electrophoresis system (see Notes 7 and 8).
3. Without UV exposure, a gel slice containing DNA fragments of 30–50 kb is then excised from the agarose, and DNA is purified from the gel using GELase and precipitated with ethanol.

4. Fosmid library construction is carried out using the CopyControl Fosmid Library Production Kit.
5. Routine sequence-based or function-based screening can be carried out depending on the aim of the study (Fig. 1). Sequence-based screening technologies are based on the known gene sequences in public databases (*e.g.* GenBank) and therefore have limited success in the finding of novel genes. Function-based approaches, on the other hand, are useful in the identification of novel functions; however, they are extremely limited by the fact that the genes from metagenomes must be expressed, and the corresponding enzymes must be folded correctly in a heterologous background. Readers are recommended to refer to recent reviews and publications for updated screening methods (1, 20, 21). Alternatively, the library can be subjected to high-throughput sequencing to reconstruct potential metabolic pathways of those micro-organisms that metabolised the labelled substrate (see Note 9).

4. Notes

1. Methods for DNA extraction from DNA-SIP incubated samples need to be selected carefully. For example, avoid methods that shear DNA (*e.g.* bead-beating protocols) when large-insert metagenomic libraries are desired. Purify the DNA before loading into CsCl gradients if humic contaminants are present, even though CsCl gradient can partially purify loaded DNA.
2. We found that phi29 DNA polymerase can be inhibited by humic substances. We recommend that DNA extracted from SIP incubations from soil samples be further purified (*e.g.* by agarose gel purification) to remove any contaminants. This will greatly enhance the yield of the amplification.
3. Typically, we found that at least 100 pg to 1 ng of DNA is required as template for MDA using the GenomiPhi V2 kit. However, lower starting template quantities (<1 ng) will yield bias in the amplification process toward DNA from certain organisms (13, 14, 18). Typically, ~4 μg of DNA will be generated from 1 ng of ^{13}C-DNA in 2 h using this kit. Longer incubations at 30°C or alternative MDA kits [reviewed in (12)] will increase this yield if necessary.
4. To assess the potential bias of phi29 DNA polymerase during MDA, we applied denaturing gradient gel electrophoresis (DGGE) to compare fingerprints of 16S rRNA gene fragments before and after MDA (13, 14). Other methods such

as microarray hybridisation may also be used. This is highly recommended since MDA is vulnerable to contamination owing to its high sensitivity.

5. During the enzyme treatment process, ~50% of the DNA may be lost. Thus, it is recommended that several MDA reactions be combined before enzyme treatment if high yields are critical. Alternatively, perform a second round of MDA using 1 μL of product from the first amplification as template. This will greatly enhance the yield of the amplification process, but potential bias introduced by this additional amplification step should also be assessed (see Note 4).
6. A standard phenol:chloroform extraction and ethanol precipitation protocol can be used here (22). In addition, we found that DNA recovery rate can be significantly increased when a phase-lock tube (*e.g.* Eppendorf) is used.
7. Settings for pulsed-field gel electrophoresis using a CHEF Mapper system (Bio-Rad) depend on the needed size of DNA fragments. To select for fragments between 30 and 50 kb, readers are recommended to refer to (14) and (13) for sample run conditions.
8. Cloning efficiency can be significantly reduced if DNA is exposed to UV light. If available, a Dark Reader transilluminator (Clare Chemical Research Inc.) should be used instead of a standard UV transilluminator. Readers are also recommended to refer to the manual of CopyControl Fosmid Library Production Kit (Epicentre) for a detailed protocol for staining and size selection of sample DNA from low-melting point agarose gel. Alternatively, DNA may be stained with Sybr stain and the gel scanned using a fluorescence scanner, such as that used for imaging 2D protein gels or DGGE fingerprints (13, 14).
9. MDA-generated DNA will contain chimeras; this is not a caveat that has been fully resolved. A recent study showed that chimeras are produced during the amplification process itself (17) and strategies to minimise chimera formation during MDA need to be further improved.

Acknowledgements

Natural Environment Research Council (NERC), the Deutsche Forschungsgemeinschaft (DFG), and the Max Planck Society are acknowledged for financial support. JDN acknowledges funding from a Discovery Grant from the National Sciences and Engineering Research Council of Canada (NSERC).

References

1. Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 68, 669–685.
2. Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., et al. (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* 5, e77.
3. Quince, C., Curtis, T.P., and Sloan, W.T. (2008) The rational exploration of microbial diversity. *ISME J* 2, 997–1006.
4. Radajewski, S., Ineson, P., Parekh, N.R., and Murrell, J.C. (2000) Stable-isotope probing as a tool in microbial ecology. *Nature* 403, 646–649.
5. Dumont, M.G. and Murrell, J.C. (2005) Stable isotope probing – linking microbial identity to function. *Nat Rev Microbiol* 3, 499–504.
6. Neufeld, J.D., Wagner, M., and Murrell, J.C. (2007) Who eats what, where and when? Isotope-labelling experiments are coming of age. *ISME J* 1, 103–110.
7. Friedrich, M.W. (2006) Stable-isotope probing of DNA: insights into the function of uncultivated microorganisms from isotopically labeled metagenomes. *Curr Opin Biotechnol* 17, 59–66.
8. Schloss, P.D. and Handelsman, J. (2003) Biotechnological prospects from metagenomics. *Curr Opin Biotechnol* 14, 303–310.
9. Wellington, E.M., Berry, A., and Krsek, M. (2003) Resolving functional diversity in relation to microbial community structure in soil: exploiting genomics and stable isotope probing. *Curr Opin Microbiol* 6, 295–301.
10. Neufeld, J.D., Dumont, M.G., Vohra, J., and Murrell, J.C. (2007) Methodological considerations for the use of stable isotope probing in microbial ecology. *Microb Ecol* 53, 435–442.
11. Dumont, M.G., Radajewski, S.M., Miguez, C.B., McDonald, I.R., and Murrell, J.C. (2006) Identification of a complete methane monooxygenase operon from soil by combining stable isotope probing and metagenomic analysis. *Environ Microbiol* 8, 1240–1250.
12. Binga, E.K., Lasken, R.S., and Neufeld, J.D. (2008) Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. *ISME J* 2, 233–241.
13. Chen, Y., Dumont, M.G., Neufeld, J.D., Bodrossy, L., Stralis-Pavese, N., McNamara, N.P., et al. (2008) Revealing the uncultivated majority: combining DNA stable-isotope probing, multiple displacement amplification and metagenomic analyses of uncultivated Methylocystis in acidic peatlands. *Environ Microbiol* 10, 2609–2622.
14. Neufeld, J.D., Chen, Y., Dumont, M.G., and Murrell, J.C. (2008) Marine methylotrophs revealed by stable-isotope probing, multiple displacement amplification and metagenomics. *Environ Microbiol* 10, 1526–1535.
15. Neufeld, J.D., Vohra, J., Dumont, M.G., Lueders, T., Manefield, M., Friedrich, M.W., et al. (2007) DNA stable-isotope probing. *Nat Protoc* 2, 860–866.
16. Chen, Y. and Murrell, J.C. (2008) DNA-Stable isotope probing: recent advances and protocols. In: Aquino de Muro, M. and Rapley, R. Ed. Methods in Molecular Biology – Gene Probes. Humana Press, Totowa, NJ.
17. Lasken, R.S. and Stockwell, T.B. (2007) Mechanism of chimera formation during the multiple displacement amplification reaction. *BMC Biotechnol* 7, 19.
18. Zhang, K., Martiny, A.C., Reppas, N.B., Barry, K.W., Malek, J., Chisholm, S.W., et al. (2006) Sequencing genomes from single cells by polymerase cloning. *Nat Biotechnol* 24, 680–686.
19. Mußmann, M., Hu, F.Z., Richter, M., de Beer, D., Preisler, A., Jorgensen, B.B., et al. (2007) Insights into the genome of large sulfur bacteria revealed by analysis of single filaments. *PLoS Biol* 5, e230.
20. Schmeisser, C., Steele, H., and Streit, W.R. (2007) Metagenomics, biotechnology with non-culturable microbes. *Appl Microbiol Biotechnol* 75, 955–962.
21. Hrvatin, S. and Piel, J. (2007) Rapid isolation of rare clones from highly complex DNA libraries by PCR analysis of liquid gel pools. *J Microbiol Methods* 68, 434–436.
22. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, New York, USA.
23. Hutchison, C.A. III and Venter, J.C. (2006) Single-cell genomics. *Nat Biotechnol* 24, 657–658.

Chapter 5

Wide Host-Range Cloning for Functional Metagenomics

Margaret Wexler and Andrew W.B. Johnston

Abstract

We describe how wide host-range cloning vectors can lead to more flexible and effective procedures to isolate novel genes by screening metagenomic libraries in a range of bacterial hosts, not just the conventionally used *Escherichia coli*. We give examples of various wide host-range plasmid, cosmid, and BAC cloning vectors and the types of genes and activities that have been successfully obtained to date. We present a detailed protocol that involves the construction and screening of a metagenomic library comprising fragments of bacterial DNA, obtained from a wastewater treatment plant and cloned in a wide host-range cosmid. We also consider future prospects and how techniques and tools can be improved.

Key words: BAC, Conjugation, Cosmid, Plasmid, IncP-1, Mobilization, Wastewater treatment, Wide host range

1. Introduction

The extent of microbial diversity in natural environments is truly astonishing. Not only are there huge numbers of species and strains, but even very closely related organisms may differ markedly in their genomic contents, due to the particular genes in their "accessory" genomes. These may form as much as half the total and may be wholly different, even in strains of the same bacterial species (1).

Furthermore, most organisms in any given environment have never been cultured in the laboratory so we know very little of their physiology, biochemistry, and other characteristics. However, the recent development of several techniques has allowed the living organisms to be bypassed, and insights into their properties have been gleaned by analyzing their genes. Two rather different ways have been used to study these "Metagenomes" – the collective genomes of all the microorganisms in an environmental sample. Both of these approaches involve harvesting DNA directly

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668,
DOI 10.1007/978-1-60761-823-2_5,

from microbes from the chosen environment, usually with no prior culturing. In one approach, this DNA is subject to sequencing on an increasingly industrial scale. This approach has been used for microorganisms obtained from (for example) soils, wastewater treatment plants, animal guts, and the oceans (see ref. 2). Such exercises generate huge data bases, which can be used to enumerate the types and numbers of different genes, and in some cases, their inferred functions. However, functions can only be attributed to genes that have very close homologues with ratified functions in known, cultured bacteria.

The second general approach involves "functional metagenomics". Here, too, gene libraries are made from environmental samples, but the cloned genes are initially examined for their ability to confer a phenotypic function to another host organism. This offers the potential of obtaining wholly novel genes whose sequences need have no similarity to those with known function (e.g., 3). Such an approach also offers the choice of prior growth of mixed cultures of the original population in quasi-natural conditions in the laboratory or mesocosm, thus enriching for the desired phenotype, such as the catabolism of a xenobiotic (e.g., 4).

By definition, screening for functionally competent metagenomic clones only detects those genes that are transcribed, translated and in which the final polypeptide can function. In most such cases, the host for the cloned DNA has been the workhorse of the molecular geneticist, *Escherichia coli*, using cloning vectors that only replicate in this host. Table 1 lists the types and properties of such vectors currently available for library construction.

They include plasmids, λ phage-based vectors, cosmids, fosmids, and bacterial artificial chromosomes (BACs). Most replicate only in *E. coli* (or its close relatives). Although such screens have been successful in several cases, this approach has limitations, some of which can be circumvented by screening the libraries in other bacterial hosts. The advantages to extending the host range of metagenomic libraries can be summarized as follows.

(a) When compared with bacterial genera such as *Pseudomonas*, *Rhizobium* and *Streptomyces*, each of which have over 15 RNA polymerase σ factors, *E. coli* only has seven, so any metagenomic genes that require specialized σ factors may not be expressed in Enteric bacteria. Indeed, only ~40% of the genes from a collection of 32 different cultured bacteria were detected by their expression in *E. coli* (9).

(b) If genes involved in complex pathways are sought (e.g., photosynthesis or secondary metabolite production), it may be impossible to clone all the genes for the entire phenotype. However, individual gene functions can be screened by using recipients with mutations in individual, corresponding genes – e.g., a photosynthetically defective mutant of *Rhodobacter*.

Table 1
Illustrative examples of vectors used for gene library construction

Vector	Host	Average insert size (kb)	Features	Advantages	Disadvantages
Plasmids, e.g., pUC18 (5)	*E. coli*	3–5	Introduce DNA by transformation or electroporation; products obtained from single genes or small operons; high copy number	Low MW DNA can be used; high plasmid copy number may help expression of certain genes even in the absence of native promoters in cloned DNA	Large genes/operons cannot be cloned; high copy number may prevent cloning of certain genes, e.g., membrane proteins. Introduction by transformation less efficient than transfection
Phage-based vectors, e.g., λZAP (6)	*E. coli*	35	λ derivatives; screen plaques	Stable, every virion contains DNA; larger operons can be cloned; efficient packaging; DNA introduced efficiently by transfection; available as component of commercial cloning kit	DNA >60 kb required; narrow host-range
Cosmids			Plasmid containing *cos* site of λ; low copy number. (usually 4–7)	Low copy number may enable cloning of certain proteins, e.g., membrane proteins; DNA introduced efficiently by transfection	DNA >60 kb required

(continued)

Table 1 (continued)

Vector	Host	Average insert size (kb)	Features	Advantages	Disadvantages
(i) Narrow host-range, e.g., pSuperCos1 (Stratagene), pWEB::TNC™ (Invitrogen)	*E. coli*	35–45	Larger operons; screen plaques or colonies, depending on vector	Large operons can be cloned; available as commercial cloning kits	Narrow host-range
(ii) Broad host range, e.g., pLAFR3 (7)	*Proteobacteria*	20–30	Store DNA as plasmid; screen colonies; introduce to other hosts by conjugation	Can be mobilized into other hosts	Not suitable for large operons
Fosmids, e.g., pCC1FOS™ (Invitrogen)	*E. coli*	35	F origin of replication; stable; Low copy number screen plaques or colonies only, depending on vector	DNA >45 kb required if blunt end cloning strategy adopted; may be induced to high copy number; introduction by transfection; available as commercial gene library cloning kit	Narrow host-range
BACs		50–100 or more	F origin of replication; stable; 1–2 copies per cell, introduce DNA by electroporation; can accommodate up to 300 kb insert DNA	Stable, suitable for very large operons	DNA >100 kb required; Extensive sub-cloning; narrow host-range; lower cloning efficiency than other vectors; size selection of DNA by pulse-field gel electrophoresis; introduce by electroporation

(i) BAC vectors, e.g., pCC1BAC™ (Invitrogen)	*E. coli*	May be induced to high copy number	Available as commercial gene library cloning kit	
(ii) BAC shuttle vectors, e.g., pMBD14 (8)	*E. coli*, *P. putida*, and *S. lividans*	Introduce library into *P. putida* or *S. lividans* by conjugation; utilizes gene replacement (*S. lividans rpsL*) to positively select for loss of plasmid	Extended host range can positively select for chromosomal integration of BACs; antibiotic resistance marker allows excised molecule to be recovered	Library must be transformed into suitable *E. coli* strain prior to transfer to *P. putida* or *S. lividans*. Specialized recipients used. BACs are integrated into chromosome, not self-replicating in non-*E. coli* host

(c) If the desired function can only be revealed in certain (usually extreme) environments (e.g., a heat-tolerant protease or other biotechnologically important enzyme), then clearly such screens are best done in a host that is itself thermotolerant, such as *Thermus*, and not *E. coli*, whose maximum temperature for growth is ~43°C.

(d) By cloning in wide host-range vectors, an inherent flexibility is built into any library. It can be "held" in a convenient bacterium (usually *E. coli*) but then can be screened repeatedly by introducing it, en masse, into other hosts for screening at a later time.

(e) A nice way to identify *cis*-acting regulatory sequences (and the corresponding catabolic genes) in metagenomes also lends itself to improvement by extending the range of hosts to screen the library. This technique, called SIGEX (Substrate-Induced Gene Expression), involves cloning metagenomic fragments 5′ of a reporter (e.g., *gfp*) and examining any clones that are activated for GFP expression by a particular environmental signal (10). These have included molecules (e.g., environmental toxins) that would activate those genes involved in their catabolism. However, SIGEX only works if the corresponding *trans*-acting regulator is present, which may not be the case in *E. coli* or any individual recipient strain.

To date, broad-host-range metagenomic libraries have been constructed using vectors based on IncP-1 plasmids or modified BACs (Table 2). The most extensively studied IncP-1 plasmids are very stable and can replicate in numerous Gram-negative bacteria. For example, the host range of IncP-1α plasmid, RK2 (17) includes Proteobacteria from the α (e.g., *Rhizobium*), β (e.g., *Thiobacillus*), δ (e.g., *Myxococcus*), and γ (e.g., *Legionella*) subgroups and photosynthetic Cyanobacteria (e.g., *Anabaena*) with *Bacteroides* being the only known Gram-negative genus that does not act as a host for RK2 (18). Most RK2-based plasmids exert only minor fitness costs to their hosts, making them suitable as vectors for genomic libraries (19). Importantly, they can also mobilize nontransmissible plasmids to many hosts including Gram-positive bacteria and even yeasts (20, 21).

Improved vectors containing (for example) polylinkers and other individual genes (e.g., *lacZ*) to detect inserts by blue/white screens have been constructed. Also, the introduction of 'phage λ *cos* sites into IncP-1 plasmids has generated cosmids, allowing larger inserts, 20–40 kb in size, to be cloned. The availability of large-insert libraries for functional metagenomics is particularly important, since genome sequencing shows that 0.2% of all prokaryotic genes are >5 kb (22). Furthermore, functions may

Table 2
Broad-host-range vectors that have been used in metagenomic library construction

Vector	Type of vector	Incompatibility group	Host range[a]	Distinguishing features	Size of vector (kb)	Average insert size (kb)	Reference
pLAFR3	Cosmid	IncP-1α	Numerous Gram-negative bacteria and some Gram positives, e.g., *Clavibacter*	Tc, RK2 replicon, RK2*oriT*, *cos*	22	16–30	7
pKS13S	Cosmid	Inc P-1α	As pLAFR3	Tc, RK2 replicon, RK2*oriT*, *cos*	21.7	16–30	4
pBBR1MCS and antibiotic resistant derivatives	Plasmid	Unknown, not Inc P-, IncQ-, IncW-, Co1E1- or P15a-based replicon	*Bordetella*, *Alcaligenes*, *Acetobacter*, *Paracoccus*, *Bartonella*, *Caulobacter*, *Rhizobium*, *Pseudomonas*, *Brucella*, *Xanthomonas*, *Rhodobacter*, *Salmonella*, *Vibrio*	Cm, RK2*oriT*, *lacZα*, *cos*	4.7	2–8	11, 12
pRK7813	Cosmid	IncP-1α	Proteobacteria	Tc, RK2 replicon, RK2*oriT*, *cos*	12.5	~33	13
pUvBBAC[b]	BAC	F replicon, Incl8	Broad host-range Gram Positive bacteria (e.g., *Listeria*, *Bacillus*, *Streptococcus*) *E. coli*	Derivative of pBeloBAC11 contains RK2*oriT*; Cm, Erm, lacZα, *cos/loxP*	11.8	68–178	14
pMBD14	BAC shuttle vector	F replicon	*E. coli*, *P. putida*, *S. lividans*	Derivative of pBeloBAC11 contains RK2*oriT* containing *Streptomyces* cassette for gene integration, Ap, Cm	~11	13–85	8

(continued)

Table 2 (continued)

Vector	Type of vector	Incompatibility group	Host range[a]	Distinguishing features	Size of vector (kb)	Average insert size (kb)	Reference
pOS7001	Cosmid shuttle vector	ColEI-based replicon	*E. coli*/*S. lividans*	Amp, Hyg, ColEI replicon, *cos* integrative into *Streptomyces* via *attP* and *int*	Not stated	50	15
pJN105	Plasmid	Same as pBBR1MCS	*E. coli*, *Bordetella*, *Vibrio*, *Pseudomonas*	*araC*-P_{Bad} cassette cloned in pBBR1MCS-5; Gm	6.1	~2–8	16
pCF430	Cosmid	Inc P-1α	as pLAFR3	pSW213 derivative; Tc RK2 replicon; RK2*oriT*, *cos*, *araC*-P_{Bad}	Not stated	Not stated	16
pRS44[b]	Wide host-range BAC/cosmid	F and Inc P-1α	as pLAFR3	Kn Cm, RK2 replicon; RK2*oriT*, *parDE*, *cos*, F factor replicon *repE*. To support plasmid replication in non-*E. coli* hosts, the transposon in pRS48 (carrying the *trfA* gene) is inserted into their chromosomes by electroporation	10.5	100–195	Patent no. WO/2007/141540

[a]In addition to *E. coli*

[b]Developed for use as broad-host-range vector for metagenomic library construction; no publications to date

require more than one gene, arranged as contiguous units in closely linked operons.

To this end, BAC vectors, which can accommodate inserts >200 kb, have been used for constructing metagenomic libraries. Normally, these only replicate in *E. coli* and contain the single-copy F-factor replicon *ori2*. However, some wide host-range shuttle BACs have been designed (Table 2), allowing their mobilization into hosts, such as *Streptomyces* and *Pseudomonas*, followed by the integration of metagenomic DNA into the host's chromosome (e.g., pMBD14, which can also be maintained in plasmid form in *Pseudomonas putida*) (8). There are disadvantages to BAC library construction, though. High molecular weight metagenomic DNA must be prepared, requiring specialized techniques such as pulse-field gel electrophoresis.

Some metagenomic studies that have employed broad-host-range or shuttle vectors (Table 3) have demonstrated directly that different hosts do indeed differentially express metagenomic DNA. Thus, Li et al. (24) showed that metagenomic *trp* genes from a sewage DNA library were differentially expressed in *E. coli* and *Rhizobium leguminosarum*, and Wexler et al. (23) showed that the same library contained a gene encoding an alcohol dehydrogenase that was expressed in *R. leguminosarum*, but not in *Pseudomonas aeruginosa* or *E. coli*. Similarly, Wang et al. (26) investigated poly-3-hydroxybutyrate metabolism genes from soil and sewage libraries and found that different classes of genes were obtained, depending on the host bacterium that was used, and Martinez et al. (8) showed that *E. coli*, *P. putida*, and *S. lividans* differ in their expression of antibiotic gene clusters when cloned into a BAC shuttle vector.

The majority of wide host-range vectors not only replicate in taxonomically diverse organisms but they can also transfer between many lineages by conjugation, involving direct cell-to-cell contact, mating-pair formation, and DNA exchange via conjugative pili (29). Because the "*tra*" genes involved in conjugal transfer are so numerous, in many cloning vectors these have been removed in order to shrink the plasmid. Thus, the widely used cosmid pLAFR3 (7) that is 22 kb in size, was derived from the 60 kb plasmid pRK2, and has lost most of the *tra* genes of its parental plasmid, so is no longer self-transmissible. However, pLAFR3 (and all other conjugally competent vectors) retains the mobilization *mob* genes, allowing them to be transferred, so long as the donor strain contains the *tra* genes, supplied either on another replicon, or integrated into the chromosome. Indeed, a widely used technique, described below, involves triparental mating, in which three strains are used, namely (a) the final recipient in which the metagenomic library will be screened for functions, (b) the culture of *E. coli* containing the pooled metagenomic library, and (c) a "helper" strain of *E. coli* that harbors a mobilizing plasmid (e.g., the widely used pRK2013) (30). This plasmid can transfer

Table 3
Examples of broad-host-range vectors used for functional metagenomic gene isolation

Function/gene target	Host[a] and relevant host phenotype	Vector	Environmental source	Screen	Library size	Average insert size (kb)	Reference
Alcohol dehydrogenase	*R. leguminosarum* 8401(pRL1) and *E. coli* 803	pLAFR3	Sewage	Growth on ethanol as sole carbon source	111,000	25	23
Tryptophan biosynthesis	*R. leguminosarum* 8401(pRL1) and *E. coli* K12 Trp-mutants	pLAFR3	Sewage	Complementation of Trp-mutants	111,000	25	24
Naphthalene dioxygenase	*P. putida* G7K2; wild-type carrying NAH7K2 (Nah-derivative of naphthalene degrading plasmid NAH7)	pKS13S	Soil	Growth on agar containing naphthalene as sole carbon source	24,000	25	4
Terragines	*S. lividans*	Not stated	Soil	HPLC-ESIMS screening	1,020	Not stated	25
Poly-D-3-hydroxybutyrate	*Sinorhizobium meliloti* Rm11107 (contains mutation in D-3-hydroxybutyrate dehydrogenase)	pRK7813	Sewage, Soil	Complementation of D-3-hydroxybutyrate dehydrogenase mutant	9,000, 34,000	33	26
Antibacterial compounds	*P. putida* MBD1 *attB* φC31 (for chromosomal integration of BAC) and *S. lividans* Δ *act red* (actinorhodin and undecylprodiginine pigments)	pMBD14	Soil	Complementation of pigmentation mutants	13,000	49	8

Polyketide synthase	*S. lividans* TK24	pOS700I	Soil	PCR	5,000	50	15
Benzoylformate decarboxylase	*P. putida* KT2440	pBBR1MCS	Soil	Growth on agar containing benzoylformate as sole carbon source	14,000	2–10	27
Antibiotic resistance	*E. coli*[b] DH10B and DH5α	pJN105	Soil	Growth on antibiotic media	200,000, 58,000, 250,000	4.1 2.7 3.5	28
		pCF430	Soil		650,000	3.5	

[a]In addition to *E. coli*

[b]Wide host-range vector was used to construct libraries; libraries were screened in *E. coli* only

by conjugation at high efficiency into the "library" culture and converts these into conjugational donors from which the wide host-range vector can transfer to the final recipient. However, pRK2013 itself has a limited host range and so cannot be maintained in the final, screening host. Finally, it is important to note that the rate of conjugal transfer is host-dependent. Thus, pLAFR3 transfers to *E. coli* and *Rhizobium* species at relatively high frequency (~10^{-1} to 10^{-2}), but transfer to other Proteobacteria, such as *Paracoccus* and *Rhodobacter*, is much less efficient (unpublished observations).

2. Materials

2.1. DNA Extraction from Wastewater

1. 2× Buffer A: 200 mM NaCl, 200 mM Tris–HCl, 2 mM sodium citrate, 10 mM calcium chloride; pH 8; autoclave, store at 4°C.
2. Polyadenylic acid potassium salt (polyA, Sigma-Aldrich), dissolved in autoclaved water at 10 mg/mL, store at −20°C.
3. Sodium pyrophosphate: prepare 10% (w/v) solution in water; filter sterilize. Store at room temperature.
4. Lysozyme: 100 mg/mL, prepare fresh with autoclaved water.
5. SDS: prepare 20% sodium dodecyl sulfate in autoclaved water. Filter-sterilize and store at room temperature.
6. Proteinase K (Qiagen): 20 mg/mL; store at 4°C.
7. Phenol:chloroform:isoamyl-alcohol (25:24:1) saturated with TE buffer, pH 8.0 (Sigma). Store at 4°C.
8. PLG tubes: Phase Lock Gel, Heavy, 50 mL tubes (Eppendorf).
9. Sodium acetate: Prepare 3 M sodium acetate, pH 5.2.
10. TE buffer: 10 mM Tris–HCl, 1 mM EDTA pH 8.0.
11. Chroma Spin + TE1000 gel filtration columns (BD Biosciences).
12. Dry ice-ethanol bath.

2.2. Partial Digestion of Wastewater DNA

1. 1× TBE running buffer: 89 mM Tris–borate, 2 mM EDTA pH 8.
2. Agarose gel: prepare in TBE buffer, add 1 μg/mL ethidium bromide while still molten.
3. 6× Loading dye: 0.25% Bromophenol blue, 30% glycerol.
4. DNA molecular weight markers: λ, λ*Hin*dIII ladder, pLAFR3.

5. Restriction enzyme *Sau*3A (e.g., Promega): dilute to 1 U/μL using enzyme dilution buffer (1× *Sau*3A buffer containing 100 μg/mL BSA).
6. BSA: 10 mg/mL acetylated bovine serum albumin (e.g., Promega).

2.3. Preparation of Broad-Host-Range Cosmid DNA

1. QIAfilter Plasmid Midi Kit (Qiagen).
2. Restriction enzyme *Bam*HI (10 U/μL).
3. Alkaline phosphatase, shrimp (Roche): 1 U/μL.
4. Ethanol: 100 and 70% in autoclaved water. Store at –20°C.

2.4. Cosmid Ligation, Packaging, Titering, and Storage

1. Gigapack® III XL Gold Packaging Extract (Stratagene, CA).
2. *E. coli* strain 803 (31) (see Note 1).
3. LB: 10 g NaCl, 10 g tryptone, 5 g yeast extract per liter, pH 7.2, sterilize by autoclaving.
4. LB agar: LB containing 1.5% agar.
5. Tetracycline stock solution: 5 mg/mL in 70% ethanol, stored at –20°C.
6. LB – Tetracycline agar: LB agar containing 5 μg/mL tetracycline, add 200 μL of tetracycline stock solution to 200 mL molten agar.
7. T_4 DNA ligase: (e.g., Promega, see Note 2).
8. 10 mM ATP (Sigma).

2.5. Conjugal Transfer of Metagenomic Libraries into Gram-Negative Host Bacteria

1. Sterile filters: sterile membrane filters without absorbent pads 0.45 μm pore size, 47 mm diameter (Whatman).
2. Kanamycin stock solution: 50 mg/mL dissolved in deionized water and filter-sterilized; store at –20°C.
3. Streptomycin stock solution: 200 mg/mL dissolved in water, filter sterilized, and stored at –20°C.
4. TY: 5 g tryptone, 3 g yeast extract, 0.9 g/L $CaCl_2 \cdot 2H_2O$ pH 6.8, sterilize by autoclaving.
5. TY agar: TY containing 1.5% agar.
6. TY-tetracycline/streptomycin agar: TY agar containing 5 μg/mL tetracycline and 400 μg/mL streptomycin. Add 200 μL of 5 mg/mL tetracycline and 400 μL of 200 mg/mL streptomycin to 200 mL molten agar.
7. Saline: prepare 0.9% (w/v) sodium chloride in deionized water. Autoclave and store at room temperature.
8. 50% glycerol solution, sterilize by autoclaving, store at room temperature.

9. *E. coli* containing helper plasmid, e.g., *E. coli* 803 (pRK2013) (30).
10. Gram-negative recipient, e.g., *R. leguminosarum* strain 3841 (32).

3. Methods

This protocol describes the isolation of wastewater DNA and the construction of a broad-host-range metagenomic library in the 22 kb cosmid vector pLAFR3, which packages fragments ~25–30 kb. The insert metagenomic DNA must be at least 60 kb to enable subsequent partial digestion prior to sticky-end ligation. The DNA extraction protocol may be adapted for use in extracting DNA from other environmental samples such as soil suspensions.

Immediately following collection of the wastewater sample, it is stored on ice and processed within 1 h. DNA may be extracted immediately or centrifuged pellets may be snap-frozen and stored at –20°C for at least several months. The protocol yields approximately 12 mg DNA per liter of wastewater. To ensure that sufficient DNA is isolated, at least two 12 mL samples should be processed.

This protocol does not include a size-fractionation step, so there is a possibility that certain clones may contain noncontiguous fragments, although the overall sizes of the inserts are ensured by using the Gigapack III XL packaging extract, in which large inserts are preferentially packaged. This means time-consuming size fractionation steps and the loss associated with sizing columns or sucrose gradients can be avoided. If large contiguous fragments are essential (e.g., if biosynthetic pathways are targeted), it is recommended that cosmid and metagenomic DNA is digested with a 6 bp-cutting restriction enzyme such as *Eco*RI. The protocol below uses *Sau*3A to digest metagenomic DNA, which is more amenable to digestion with this enzyme.

3.1. DNA Extraction from Wastewater

1. Centrifuge 12 mL wastewater at 4,500 × *g* for 20 min at 4°C; decant supernatant (see Note 3).
2. Resuspend pellet in 12 mL 2× Buffer A; repeat centrifugation step.
3. Add 6 mL 2× Buffer A, 240 μL polyA, 240 μL 10% pyrophosphate, and 360 μL lysozyme in that order. Mix by gentle inversion. Incubate for 40 min at 37°C.
4. Add 120 μL 20% SDS and 720 μL proteinase K. Mix by gentle inversion. Incubate for 30 min at 50°C.

5. Add 2.4 mL 20% SDS. Carry out three cycles of freeze–thaw by placing tubes in dry-ice ethanol bath (–70°C) for 3 min, then in a 65°C water bath for 5 min.
6. Immediately prior to use, centrifuge PLG tubes at 1,500 × *g*, 4°C, for 1–2 min (see Note 4).
7. Transfer lysate (from step 5) to 50 mL PLG tube.
8. Add equal volume of phenol–chloroform to tube. Mix by gentle inversion for approximately 5 min.
9. Centrifuge for 5 min at 1,500 × *g* to separate phases.
10. Carefully decant the upper, aqueous layer (which contains nucleic acids) to a fresh PLG tube.
11. Repeat steps 6–9 at least three times until there is no protein or chromosomal debris visible at the gel/aqueous interphase.
12. Transfer aqueous phase acid to 2-mL microfuge tubes (1 mL of sample per tube). Precipitate nucleic acids by adding 110 μL of 3 M sodium acetate, pH 5.2 and 800 μL of propan-2-ol. Leave tubes on ice for 20 min, then centrifuge at 12,000 × *g* for 20 min at 4°C (see Note 5).
13. Decant supernatant. Rinse pellets in 1.9 mL cold 70% ethanol. Air-dry DNA for ~30 min. Add 25 μL TE per microfuge tube. Heat in a 70°C waterbath for 10 min, then leave at 4°C overnight. Very gently resuspend any remaining undissolved DNA by gentle mixing. If DNA is difficult to dissolve, add a further 25 μL TE buffer. Do not pipette or vortex.
14. Purify the DNA using Chroma Spin gel filtration columns according to manufacturer's instructions. A maximum volume of 100 μL can be purified per column.
15. Store DNA at –20°C.

3.2. Partial Digestion of Wastewater DNA

1. Estimate concentration of wastewater DNA by spectrophotometry using an OD_{260}. The OD_{260}/OD_{280} ratio should be 1.7–1.9 and the protein concentration <0.1 mg/mL.
2. Run uncut wastewater DNA (approximately 100, 200, and 500 ng) on a 0.3% agarose gel at approximately 30 V for 48 h at 4°C to estimate DNA size and concentration. Include known amounts (50, 100, 200, and 300 ng) of uncut λ, 1 μg of λ*Hin*dIII ladder and 100 ng of pLAFR3 (see Subheading 2.2) as size standards.
3. Test conditions for partial digestion: Add the following to a 1.5-mL microfuge tube: 10 μg wastewater DNA, 10 μL 10× *Sau*3A enzyme buffer, 1 μL of BSA (10 mg/mL) to a final volume of 99 μL. Prewarm microfuge tube to 37°C. To begin digests, add 1 μL *Sau*3A (1 U/μL). After 30 s, remove 10 μL of digest and immediately transfer to microfuge tube containing 4 μL loading dye. Continue to

remove 10 μL aliquots after 1, 2, 3, 4, 5, 6, 7, and 10 min. Incubate the remaining digest for a further hour, then add 4-μL loading dye.

4. Run partially digested DNA on a 0.7% agarose gel. Include as standards: λ, λ*Hin*dIII (see Subheading 3.2, step 2) and uncut wastewater DNA (1 μg).
5. Determine which sample contains most of its fragments between the sizes of uncut λ (48 kb) and the largest λ*Hin*dIII fragment (23 kb). If samples are overdigested, dilute *Sau*3A to 0.5 U/μL and repeat steps 3–5. If samples are underdigested, repeat with increased incubation times.
6. Choose two digestion times in which the majority of fragments are around 35 kb. Repeat partial digestion reaction for these two time periods (Subheading 3.2, step 3).
7. Once reaction is complete, immediately transfer sample to prepared Chromo Spin gel filtration column and purify DNA (see Subheading 3.1, steps 14 and 15, see Note 6).
8. Determine DNA concentration and size as previously (see Subheading 3.2, steps 1 and 4). If most of the digested DNA fragments are not the appropriate size, repeat partial digestion reactions using a shorter (or longer) incubation time.

3.3. Preparation of Broad-Host-Range Cosmid DNA

1. Prepare cosmid DNA using QIAfilter Plasmid Midi Kit according to manufacturer's instructions.
2. Resuspend DNA in TE buffer. Measure OD_{260}. Ensure DNA concentration is at least 100 μg/mL. Ensure protein concentration is <0.1 mg/mL and OD_{260}/OD_{280} ratio is between 1.7 and 1.9.
3. Check DNA concentration by running 100 and 200 ng cosmid DNA on a 0.7% agarose gel; use known amounts (50, 100, 200, and 300 ng) of uncut λ as standards.
4. Digest 25 μg DNA with 5 μL *Bam*HI (50 U) in a 250 μL. Leave at 37°C for 3 h. Inactivate enzyme by heating to 65°C for 15 min.
5. Run 200 ng each of digested and undigested DNA sample on a 0.7% agarose gel to determine if sample is fully digested. Include 750 ng of λ*Hin*dIII ladder as size standard.
6. Ethanol precipitate DNA with one of ten volume sodium acetate (pH 5.2) and 2 volumes ethanol. Incubate at –20°C for at least 30 min. Centrifuge for 15 min, wash twice in 1 mL 70% ethanol, dry pellet, and resuspend in 30 μL TE buffer. Determine DNA concentration by agarose gel electrophoresis and spectrophotometry (Subheading 3.3, steps 2 and 3).
7. Incubate 5 μg of *Bam*HI-digested cosmid DNA with 1 U shrimp alkaline phosphatase in 50 μL, at 37°C for 30 min.

Add 1 U alkaline phosphatase and incubate a further 30 min. Inactivate enzyme by heating to 65°C for 15 min.

8. Extract DNA with an equal volume of phenol:chloroform isoamyl alcohol; mix gently for 5 min, then spin for 10 min. Carefully remove aqueous layer then ethanol precipitate DNA (as in Subheading 3.3, step 6) and resuspend in 20 μL of TE buffer.
9. Determine DNA concentration by agarose gel electrophoresis and spectrophotometry, as in Subheading 3.3, steps 2 and 3 (see Note 7).

3.4. Cosmid Ligation, Packaging, Titering, and Storage

1. Ligation: Add 1 μg *Bam*HI-digested dephosphorylated pLAFR3, 2.5 μg *Sau*3A partially digested wastewater DNA, 1.5 μL T_4 DNA ligase, and 1 mM ATP to a final volume of 20 μL. Incubate at 14°C overnight (see Note 8).
2. Package 1.5 μL ligation reaction (~0.26 μg) using Gigapack III XL Gold Packaging Extract according to manufacturer's instructions.
3. Prepare host bacteria, e.g., *E. coli* 803, titer and store library in 50% glycerol in 2-μL microfuge tubes, according to Stratagene's instructions.
4. To ensure cosmid library contains inserts containing metagenomic DNA, purify, for example, 12 colonies on LB-tetracycline agar plates, extract, cosmid DNA digest, run on agarose gels.

3.5. Conjugal Transfer of Metagenomic Libraries into Gram-Negative Host Bacteria

1. Inoculate recipient (e.g., *R. leguminosarum*) into 5 mL TY. Incubate with shaking at 28°C overnight.
2. Inoculate helper plasmid, e.g., *E. coli* 803(pRK2013), into 5 mL LB broth containing 20 μg/mL kanamycin. Incubate with shaking at 37°C overnight.
3. Transfer aliquot (20–100 μL) of *E. coli* 803 containing sample of entire metagenomic library to 5 mL LB containing 5 μg/mL tetracycline. Incubate with shaking at 37°C overnight.
4. Spin down 3 mL of each bacterial culture and wash in 3 mL 0.9% sterile saline.
5. Resuspend in 300 μL TY.
6. Filter crosses: aseptically transfer a sterile filter to a TY agar plate. Add 100 μL of each strain to the filter. Leave on bench to dry. Incubate at 28°C overnight.
7. Controls: add two of each of the above strains to a filter, as in step 6. Leave on bench to dry. Incubate at 28°C overnight.

8. Transfer filter containing bacteria to a sterile universal tube. Wash off cells with TY. Prepare cell dilutions 10^{-2}, 10^{-4}, and 10^{-6} using 0.9% saline. Plate 100 μL of each dilution onto TY-streptomycin/tetracycline agar.
9. Plate remaining triparental cross (undiluted) onto TY-streptomycin/tetracycline agar plates, 100 μL per plate and incubate at 28°C for 3–4 days.
10. Wash cells from TY-streptomycin/tetracycline agar with undiluted cells by adding 0.5 mL TY broth containing streptomycin (200 μg/mL) and tetracycline (5 μg/mL). Add equal volume of 50% glycerol. Store at –20°C in 50% glycerol in 1 mL aliquots.

3.6. Future Prospects

The limited number of studies on wide host-range metagenomic libraries has shown that these vectors have a role to play in the exploitation of functional metagenomics to identify novel gene products. However, there is a pressing need to develop better, more sophisticated vectors that overcome some of the problems that are associated with those that have been used to date. For example, pLAFR3 can only accommodate ~25–30 kb, being 22 kb in size, and has a copy number of only 4–7 (30), making it more difficult to isolate than smaller vectors.

An example of the way ahead for designer vectors for metagenomic library construction is the recently patented pRS44 (Patent WO/2007/141540). This RK2-based broad-host-range cloning vector is a modified BAC that includes *ori2* and *repE*, *cosN* (for packaging in λ) the RK2 origins of replication (*oriV*) and conjugal transfer (*oriT*) and *parDE*. It allows stable maintenance of large (>30 kb) inserts at copy numbers (either high or medium) that can be manipulated by the experimenter.

Metagenomics is a relatively new concept, requiring new approaches, and it is clear that no one method can or should be used to exploit or to understand the remarkable diversity that is all around us, wherever microbes gather together "out there" in their chosen environments. However, it seems likely that functional metagenomics will contribute to these analyses and that, as the technology improves, it will have an increasingly important role, particularly if it encompasses the use of wide host-range screening systems to increase the flexibility of this approach.

4. Notes

1. *E. coli* strain 803 is particularly suitable for the transformation of large plasmids.
2. Ligase buffer must not contain polyethylene glycol, which can inhibit the packaging reaction.

3. Take care when handling wastewater DNA, since it may contain pathogenic microorganisms and viruses. Use mask and gloves and autoclave waste tips and tubes immediately after use.
4. The use of PLG tubes results in enhanced recovery of DNA by providing a barrier between the organic and aqueous phases.
5. Use wide-bore or cut off pipette tips to avoid shearing DNA.
6. The use of Chromo Spin gel filtration columns for purifying partially digested metagenomic DNA has the additional advantage of removing small DNA fragments.
7. Success of phosphatase treatment of cosmid DNA can be determined by performing self-ligation reactions and ligating fragments of known size to the cosmid, using standard molecular biology techniques.
8. Ligation reactions should be carried out at DNA concentrations of 0.2 μg/μL or greater, which favors concatemers and not circular DNA molecules that have only one *cos* site.

References

1. Binnewies, T.T., Motro, Y., Hallin, P.F., Lund, O., Dunn, D., La, T., et al. (2006) Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. *Funct Integr Genomics* 6, 165–168. *http://www.ncbi.nlm.nih.gov/genomes/lenvs.cgi*.
2. *http://www.ncbi.nlm.nih.gov/genomes/lenvs.cgi*
3. Suenaga, H., Ohnuki, T., and Miyazaki, K. (2007) Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environ Microbiol* 9, 2289–2297.
4. Ono, A., Miyazaki, R., Sota, M., Ohtsubo, Y., Nagata, Y., and Tsuda, M. (2007) Isolation and characterization of naphthalene-catabolic genes and plasmids from oil-contaminated soil by using two cultivation-independent approaches. *Appl Microbiol Biotechnol* 74, 501–510.
5. Yanisch-Perron, C., Vieira, J., and Messing, J., (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33, 103–119.
6. Short, J.M., Fernandez, J.M., Sorge, J.A., and Huse, W.D. (1988) Lambda ZAP: a bacteriophage lambda expression vector with *in vivo* excision properties. *Nucleic Acids Res* 16, 7583–7600.
7. Staskawicz, B., Dahlbeck, D., Keen, N., and Napoli, C. (1987) Molecular characterization of cloned avirulence genes from race 0 and race 1 of *Pseudomonas syringae* pv. *glycinea*. *J Bacteriol* 169, 5789–5794.
8. Martinez, A., Kolvek, S.J., Yip, C.L., Hopke, J., Brown, K.A., MacNeil, I.A., et al. (2004) Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Appl Environ Microbiol.* 70, 2452–2463.
9. Gabor, E.M., Alkema, W.B.L., and Janssen, D.B. (2004) Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ Microbiol* 6, 879–886.
10. Uchiyama, T., Abe, T., Ikemura, T., and Watanabe, K. (2005) Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat Biotechnol* 23, 88–93.
11. Kovach, M.E., Phillips, R.W., Elzer, P.H., Roop, II, R.M., and Peterson, K.M. (1994) pBBR1MCS: a broad-host-range cloning vector. *Biotechniques* 16, 800–802.
12. Kovach, M.E., Elzer, P.H., Hill, D.S., Robertson, G.T., Farris, M.A., Roop, II, R.M., et al. (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* 166, 175–176.
13. Jones, J.G.D. and Gutterson, N. (1987) An efficient mobilizable cosmid vector, pRK7813,

and its use in a rapid method for marker exchange in *Pseudomonas fluorescens* strain HV37a. *Gene* 61, 299–306.

14. Hain, T., Otten, S., von Both, U., Chatterjee, S.S., Technow, U., Billion, A., et al. (2008) Novel bacterial artificial chromosome vector pUvBBAC for use in studies of the functional genomics of *Listeria* spp. *Appl Environ Microbiol* 74, 1892–1901.
15. Courtois, S., Cappellano, C.M., Ball, M., Francou, F.X., Normand, P., Helynck, G. et al (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl Environ Microbiol* 69, 49–55.
16. Newman, J.R. and Fuqua, C. (1999) Broad-host-range expression vectors that carry the L-arabinose-inducible *Escherichia coli ara*BAD promoter and the *araC* regulator. *Gene* 227, 197–203.
17. Thomas, C.M. and Smith, C.A. (1987) Incompatibility group P plasmids: genetics, evolution, and use in genetic manipulation. *Annu Rev Microbiol* 41, 77–101.
18. Thomas, C.M. and Helinski, D.R. (1989) Vegetative replication and stable inheritance of IncP plasmids. In: Thomas, C.M. Ed. Promiscuous Plasmids of Gram-Negative Bacteria. Academic, San Diego, pp. 1–25.
19. De Gelder, L., Ponciano, J.M., Joyce, P., and Top, E.M. (2007) Stability of a promiscuous plasmid in different hosts: no guarantee for a long-term relationship. *Microbiology* 153, 452–463.
20. Schlüter, A., Szczepanowski, A., Pühler, A., and Top, E.M. (2007) Genomics of IncP-1 antibiotic resistance plasmids isolated from wastewater treatment plants provides evidence for a widely accessible drug resistance gene pool. *FEMS Microbiol Rev* 31, 449–477.
21. Adamczyk, M. and Jagura-Burdzy, G. (2003) Spread and survival of promiscuous IncP-1 plasmids. *Acta Biochim Pol* 50, 425–453.
22. Reva, O.N. and Tũmmler, B. (2008) Think big–giant genes in bacteria. *Environ Microbiol.* 10, 768–777.
23. Wexler, M., Bond, P.L., Richardson, D.J., and Johnston, A.W.B. (2005) A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. *Environ Microbiol* 7, 1917–1926.
24. Li, Y.G., Wexler, M., Richardson, D.J., Bond, P.L., and Johnston, A.W.B. (2005) Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of *Rhizobium leguminosarum* and of *Escherichia coli* reveals different classes of cloned *trp* genes. *Environ Microbiol* 7, 1927–1936.
25. Wang, G.-Y.-S., Graziani, E., Waters, B., Pan, W., Li, X., McDermott, J., et al. (2000). Novel natural products from soil DNA libraries in a streptomycete host. *Org Lett* 2, 2401–2404.
26. Wang, C., Meek, D.J., Panchal, P., Boruvka, N., Archibald, F.S., Driscoll, B.T., et al. (2006) Isolation of Poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants. *Appl Environ Microbiol* 72, 384–391.
27. Henning, H., Leggewie, C., Pohl, M., Müller, M., Eggert, T., and Jaeger, K.-E. (2006) Identification of novel benzoylformate decarboxylases by growth selection. *Appl Environ Microbiol* 72, 7510–7517.
28. Riesenfeld, C.S., Goodman, R.M., and Handelsman, J. (2004) Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ Microbiol* 6, 981–989.
29. Sørensen, S.J., Bailey, M., Hansen, L.H., Kroer, N., and Wuertz, S. (2005) Studying plasmid horizontal transfer *in situ*: a critical review. *Nat Rev Microbiol* 3, 700–710.
30. Figurski, D.H. and Helinski, D.R. (1979) Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc Natl Acad Sci USA* 76, 1648–1652.
31. Wood, W.B. (1966) Host specificity of DNA produced by *Escherichia coli;* bacterial mutations affecting the restriction and modification of DNA. *J Mol Biol* 16, 118–133.
32. Young, P., Crossman, L.C., Johnston, A.W.B., Thomson, N.R., Ghazoui, Z.F., Hull, K.H., et al. (2006) The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol* 7, R34.

Chapter 6

Cloning and Expression Vectors for a Gram-Positive Host, *Streptomyces lividans*

Kristof Vrancken, Lieve Van Mellaert, and Jozef Anné

Abstract

The choice of an expression system for the meta-genomic DNA of interest is of vital importance for the detection of any particular gene or gene cluster. Most of the screens to date have used the Gram-negative bacterium *Escherichia coli* as a host for the meta-genomic gene libraries. However, the use of *E. coli* introduces a potential host bias since only 40% of the enzymatic activities may be readily recovered by random cloning in *E. coli* (Gabor et al., Environ Microbiol 6:879–886, 2004). To recover some of the remaining 60%, alternative cloning hosts such as *Streptomyces* spp. have been used (Lorenz and Eck, Nat Rev Microbiol 3:510–516, 2005). *Streptomycetes* are high-GC Gram-positive bacteria that belong to the *Actinomycetales*, and they have been studied extensively in the last 10 years as an alternative expression system (reviewed in Vrancken and Anné, Future Microbiol 4:181–188, 2009). *Streptomyces* is extremely well suited for the expression of DNA from other actinomycetes and genomes of high GC content (Wang et al., Org Lett 2:2401–2404, 2000). Furthermore, due to its high innate secretion capacity, it can be a superior system than *E. coli* for the production of many extra-cellular proteins.

Key words: *Streptomyces*, Expression, Cloning, Actinomycetes, Secretion

1. Introduction

Streptomyces lividans has, in recent years, shown to be a promising expression system for several proteins that are difficult to be expressed in other bacterial host systems such as *Escherichia coli* (reviewed in ref. 3). Although a broad range of genes, both pro- and eukaryotic, has been expressed in this host, *S. lividans* is particularly useful for the expression of genes from actinomycetes and genes from other genomes of high GC content. Given the abundance of actinomycetes in soil samples, it comes as no surprise that the first expression of meta-genomic DNA in *Streptomyces* was of a soil sample (4) and resulted in the discovery of novel bioactive

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668, DOI 10.1007/978-1-60761-823-2_6,

compounds. The recent development of *E. coli–S. lividans* cosmid shuttle vectors (5) greatly facilitates the expression of entire meta-genomic libraries, since they allow the construction of the libraries in the standard host, *E. coli*, while the screening can be performed in either *E. coli* or *S. lividans*. In a next step, enzymes of interest can easily be produced in *S. lividans* due to the availability of a wide range of expression vectors. Novel enzymes, identified in soil and marine meta-genomic screens, have already been produced to high levels in *S. lividans* (6, 7).

In this chapter, we discuss all the protocols necessary for researchers to grow *S. lividans* TK24, the preferred host thanks to the absence of a methylation/restriction system and a low protease activity, to express entire meta-genomic libraries and finally to express single genes of interest in this host.

It is important to note that while this chapter discusses mostly *S. lividans* TK24, most of these protocols are readily applicable to other *Streptomyces* spp. and where differences occur, this is mentioned in Subheading 4.

2. Materials

2.1. Growth of *S. lividans*

1. Phage medium: 0.5 g $MgSO_4{\cdot}7H_2O$, 0.74 g $CaCl_2{\cdot}2H_2O$, 10 g glucose, 5 g tryptone (Becton–Dickinson, cat. no. 211705), 5 g yeast extract (Becton–Dickinson, cat. no. 288620), 5 g Lab Lemco powder (Oxoid, cat. no. LP0029B). Bring to 1 L with deionised water (dH_2O). Adjust the pH of the solution to 7.2 with 5 N NaOH and sterilise.
2. Spore element solution: 40 mg $ZnCl_2$, 200 mg $FeCl_3{\cdot}6H_2O$, 10 mg $CuCl_2{\cdot}2H_2O$, 10 mg $MnCl_2{\cdot}4H_2O$, 10 mg $Na_2B_4O_7{\cdot}10H_2O$, 10 mg $(NH_4)_6Mo_7O_{24}{\cdot}4H_2O$. Bring to 1 L with dH_2O and filter-sterilise.
3. TES buffer: 0.25 M TES, pH 7.2.
4. R2 medium: Dissolve 103 g sucrose, 0.25 g K_2SO_4, 12.12 g $MgCl_2{\cdot}6H_2O$, 0.1 g casamino acids (Becton–Dickinson, cat. no. 223050), 1 g yeast extract (Becton–Dickinson, cat. no. 288620), 5 g of Lab Lemco powder (Oxoid, cat. no. LP0029B). Add 100 mL TES buffer, 2 mL of the spore element solution and 10 mL of a KH_2PO_4 (0.5%) solution. Bring to 1 L with dH_2O. Divide the suspension in 4×250 mL Erlenmeyer and add 5.5 g of agar to each Erlenmeyer. Autoclave for 20 min. Add 1/100 volumes of a filter-sterile 36.8% $CaCl_2{\cdot}2H_2O$ and 1/1,000 volumes of a filter-sterile 2 mM $CuSO_4$ solution (see Notes 1 and 2) and pour into Petri dishes.

5. Glass/Teflon Elvehjem-Potter cell homogeniser (see Note 3 and Fig. 1).
6. TSB medium: 30 g of Tryptone Soya Broth powder (Oxoid, cat. no. CM129). Bring to 1 L with dH_2O (see Note 4).

2.2. Preparation of *S. lividans* Spore Suspension

1. 20% glycerol in dH_2O (sterile).
2. Three- to four-day-old culture of *S. lividans* plated on MS medium.
3. Inoculating loop.
4. Sterilised 10-mL syringes containing non-absorbent cotton wool (see Note 5 and Fig. 2).
5. 12-mL Falcon tube.

2.3. Plasmid Conjugation from *E. coli* to *S. lividans*

1. *E. coli* S17-1 (ATCC #4705) or *E. coli* ET12567 [pUZ8002] (8) cells containing the DNA of interest (see Note 6).
2. Lysogeny broth (LB): 10 g tryptone (Becton–Dickinson, cat. no. 211705), 5 g yeast extract (Becton–Dickinson, cat. no. 288620), 10 g NaCl. Bring to 1 L with deionised water (dH_2O). Adjust to pH 7.0 with 5 N NaOH and sterilise.
3. 2× YT medium: 32 g tryptone (Becton–Dickinson, cat. no. 211705), 20 g yeast extract (Becton–Dickinson, cat. no. 288620), 10 g NaCl. Bring to 1 L with dH_2O water and sterilise.
4. Mannitol soya flour (MS) medium: Dissolve 20 g of mannitol in 1 L of tap water. Add 20 g of agar and 20 g of soya flour (see Note 7) to the solution. *Autoclave twice* with gently shaking between both runs. Add 10 mM $MgCl_2$ and pour into plastic Petri dishes.
5. Antibiotic stock solutions (where appropriate): Ampicillin (50 mg/mL in dH_2O), apramycin (50 mg/mL in dH_2O), kanamycin (50 mg/mL in dH_2O), nalidixic acid (25 mg/mL in 0.2 N NaOH), thiostrepton (50 mg/mL in DMSO).

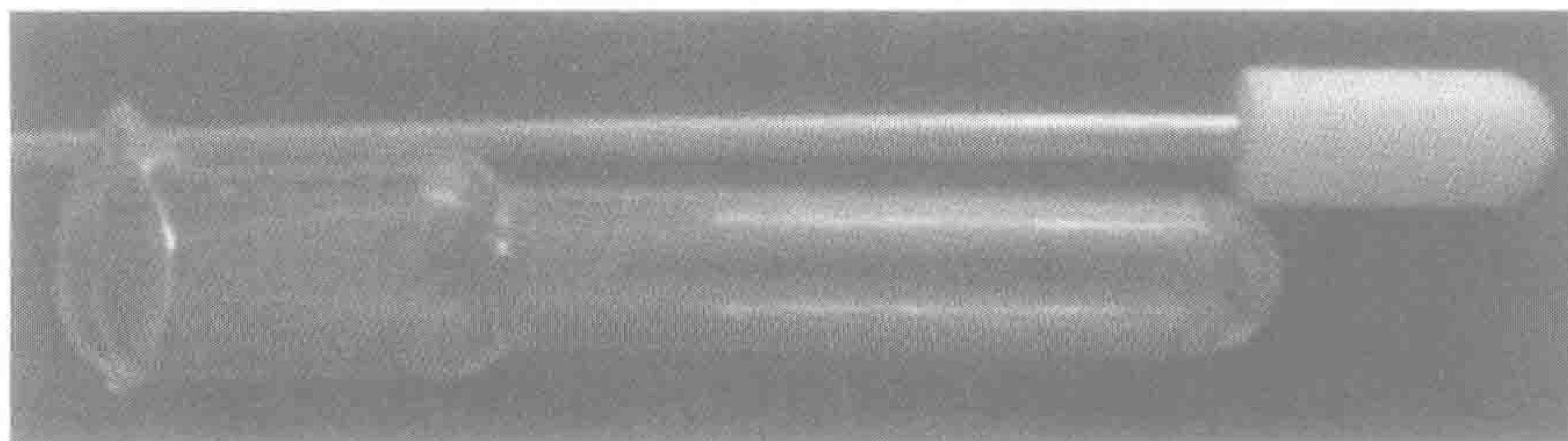


Fig. 1. Potter cell homogeniser.

2.4. Preparation of Streptomyces spp. Protoplasts

1. Phage medium (see Subheading 2.1).
2. 6.5% glucose in dH_2O, filter sterilised.
3. 20% glycine in dH_2O, autoclaved.
4. S-medium: Dissolve in 800 mL dH_2O: 4 g peptone (Becton–Dickinson, cat. no. 211921), 4 g yeast extract (Becton–Dickinson, cat. no. 288620), 0.5 g $MgSO_4 \cdot 7H_2O$, 2 g KH_2PO_4 and 4 g K_2HPO_4. Divide in 3× 266 mL and autoclave for 20 min. Add 50 mL of 6.5% glucose solution and 0.8% glycine (final concentration) to 266 mL medium (see Note 8).
5. A pre-culture of *S. lividans* in 5 mL phage medium, grown at 300 rpm for 48 h.
6. A sterile 0.9% NaCl solution.
7. Spore element solution (see Subheading 2.1).
8. TES buffer (see Subheading 2.1).
9. PTC buffer: 103 g sucrose, 0.25 g K_2SO_4, 2.03 g $MgCl_2 \cdot 6H_2O$, 2.94 g $CaCl_2 \cdot 2H_2O$, 80 mL TES buffer, 2 mL spore element solution. Bring to 1 L with dH_2O and autoclave.
10. Lysozyme (Roche diagnostics).

2.5. Protoplast Transformation

1. PTC buffer (see Subheading 2.4).
2. Filter-sterilised solution of 35% PEG6000 (NBS Biologicals, cat. no. 14808-C) in PTC buffer (see Note 9).
3. R2 medium (see Subheading 2.1).
4. Stock solutions of the appropriate antibiotics.

3. Methods

3.1. Growth of S. lividans

S. lividans and *Streptomyces* spp., in general, are relatively easy to grow. *S. lividans* grows much slower than *E. coli* though, and a 5-mL culture can take 48 h to grow to a sufficient density for further experiments. Contrary to *E. coli, S. lividans* does not show fully dispersed growth but tends to grow as pellets of mycelium. These pellets can be troublesome to work with, especially when using a culture to inoculate a second one. However, mechanical homogenisation, or the addition of dispersants to the medium can greatly reduce this pelleted growth. Here, we describe a basic workflow to grow *Streptomyces* spp., starting from a colony obtained as a spore suspension or glycerol stock and leading to a 50-mL flask culture.

1. Pour 20 mL of R2 medium into a standard Petri plate and spread 100 μL of the spore suspension or glycerol stock on this medium using a glass spreader (see Note 10).
2. After 2–3 days, use an inoculation loop to pick off a single colony and re-suspend this colony in 5-mL phage medium. This culture can be used as a starter for the 50-mL flask culture (see Note 11).
3. Incubate at 27 °C with shaking at 300 rpm for 48–60 h.
4. Pour the culture in a glass cell homogeniser and move the Teflon piston up and down in the suspension to homogenise the mycelium pellets.
5. Pipette 1 mL of this homogenised culture in an Erlenmeyer flask containing 50 mL of TSB medium (see Note 12) and incubate this culture at 27 °C at 300 rpm. After 24–48 h of growth, this culture can be used for further analysis (e.g. enzymatic activity, secondary metabolites).

3.2. Preparation of S. lividans Spore Suspension

Streptomyces spore suspensions are a very useful tool. They are easier to work with than standard glycerol stocks (20% final glycerol concentration), and inoculation with spore suspension usually results in cultures that are quicker in reaching the required density for further experiments (DNA/RNA isolation, enzyme assays, etc.). Furthermore, conjugation to spores is more efficient than conjugation to mycelial fragments. Therefore, spore suspensions are essential when conjugating a meta-genomic DNA library from *E. coli* to *S. lividans*.

1. Pour four MS plates, adding any necessary antibiotics to the plate.
2. Spread 1 mL of an overnight culture of *S. lividans* grown in phage medium on each plate.
3. Incubate the plates for 4–5 days at 30 °C (see Note 13).
4. Add 9 mL of sterile dH_2O to the plate.
5. Use an inoculation loop or sterile cotton bud to harvest the spores by gently scraping the surface of the culture. Gradually increase the pressure on the surface and scrape more vigorously, without damaging the agar.
6. Pipette or pour the spore suspension into a 12-mL Falcon tube and vortex the suspension at maximum setting to break the spore chains.
7. Bring the suspension through a sterile syringe containing non-absorbent cotton wool (see Note 5 and Fig. 2) into a 12-mL Falcon tube.
8. Spin the Falcon tube at $5,000 \times g$ for 5 min and immediately pour off the supernatant.



Fig. 2. Syringe with cotton wool, used to filter the spore suspension.

9. Re-suspend the spores in 1–2 mL of sterile 20% glycerol and vortex briefly. Freeze at –20 °C.
10. It is often desirable to have a rough idea of the amount of spores. Therefore, take 1 μL of the spore suspension and use this to make tenfold dilutions (down to 10^{-9}) in dH_2O. Plate these dilutions on MS medium and incubate them at 27–30 °C for 2–3 days after which a CFU count can be done.

3.3. Plasmid/Cosmid Conjugation from *E. coli* to *S. lividans*

Introducing DNA into *S. lividans* can be done either by protoplast transformation or by conjugation. The latter has several distinct advantages, the main one being that the vectors can be replicated in *E. coli*, greatly facilitating the production of the required constructs. Furthermore, protoplast transformation is very inefficient when larger DNA fragments such as cosmids are introduced, which have very little influence on the conjugation efficiency. *E. coli–S. lividans* cosmid shuttle vectors are available (4, 5, 9), and these vectors allow the construction of libraries using the standard host *E. coli*, but the subsequent screening can be performed employing *E. coli* or *S. lividans* as hosts. Here, we describe a standard protocol to perform the conjugation of cosmids from *E. coli* to *S. lividans*.

1. Transform competent *E. coli* S17-1 (ATCC #4705) or *E. coli* ET12567 [pUZ8002] cells containing the DNA of interest with the *oriT*-containing cosmid (see Note 6).
2. Resuspend one colony into 5 mL of LB medium supplemented with the appropriate antibiotic(s) to select for the *oriT*-containing plasmid and grow overnight at 30 °C (see Note 14).
3. Dilute the overnight culture 1:100 in fresh LB medium and grow at 37°C to an OD_{600} of 0.4–0.5.
4. Centrifuge the cells at 5,000 × *g* for 5 min (see Note 15).
5. Decant the supernatant and re-suspend the cell pellet in an equal volume of ice cold LB.
6. Repeat steps 4–5–4 in this order.
7. Finally, re-suspend the cell pellet in 0.1 volume of ice-cold LB and place the suspension on ice.

8. While washing the *E. coli* cells, add 10^8 spores to 0.5 mL of 2× YT medium.
9. Centrifuge the spores at 13,000 × g for 1 min.
10. Decant the supernatant and re-suspend the spores in 0.5 mL of 2× YT medium.
11. Repeat steps 9–10–9–10 in this order.
12. Use a heat block to incubate the spore mix at 59 °C for 10 min and then allow the mixture to cool to room temperature.
13. Add 500 μL of the *E. coli* suspension to the spore mixture. Vortex and spin briefly.
14. Pour off the supernatant and re-suspend the pellet in the remaining fluid.
15. Plate on MS medium supplemented with 10 mM $MgCl_2$ (see Note 16) and incubate the plates at 27–30 °C for 16–20 h.
16. Overlay the plate with 1 mL dH_2O containing 0.5 mg nalidixic acid and the appropriate antibiotic to select for successful exconjugants (see Note 17).
17. Spread the antibiotic solution evenly (see Note 18).
18. Continue incubation at 27–30 °C for 3–4 more days.
19. Pick off potential exconjugants to selective media containing 25 μg/mL nalidixic acid.

3.4. Preparation of *Streptomyces* spp. Protoplasts

Once an enzymatic activity of interest (or a bioactive compound) has been identified in an *S. lividans* library, it might be desirable to express a single gene instead of an entire genome region. Several vector and expression systems are currently available for *S. lividans*, with different advantages and disadvantages (3, 10). The expression cassette can be constructed either in shuttle vectors, which can replicate in both *E. coli* and *S. lividans* or directly in an *S. lividans* vector. In the former case, *S. lividans* will have to be transformed with a purified plasmid, while in the latter case they have to be transformed with a ligation mixture. Transformation of *S. lividans* cells is done using protoplasts. *S. lividans* protoplasts can readily be transformed by plasmid DNA at very high frequency in the presence of PEG 6000. In the following two paragraphs, the protocol for the preparation and transformation of these protoplasts is discussed.

1. Pre-culture *S. lividans* in 5 mL phage medium for 48 h. If necessary, add appropriate antibiotic.
2. Homogenise culture (as described in Subheading 3.1) and inoculate 50 mL S-medium with 2 mL preculture. Incubate this culture for 20–24 h at 27–30 °C at 280 rpm.
3. Harvest the culture by centrifugation at 5,000 × g for 5 min.

4. Decant the supernatant and re-suspend the cells in 0.9% NaCl.
5. Centrifuge the cells at 5,000 × *g* for 5 min.
6. Carefully decant the supernatant and re-suspend the cells in 15 mL PTC buffer.
7. Centrifuge the cells at 5,000 × *g* for 5 min.
8. During this centrifugation step, prepare, per sample, 5.5 mL of PTC buffer containing 10 mg/mL lysozyme and filter-sterilise this solution.
9. Re-suspend the cell pellet in 5 mL of this lysozyme solution and incubate the cell suspension at 27–30 °C on a rotary shaker (120 rpm) for 15–30 min (see Note 19).
10. Check the formation of protoplasts using a phase-contrast light microscope (see Note 20).
11. If sufficient protoplasts are formed continue to step 12, otherwise prolong the incubation in the lysozyme solution.
12. Add 10 mL PTC buffer, gently pipette the suspension up and down and centrifuge the suspension at 800 × *g*. This will leave the protoplasts in suspension, while the mycelium fragments are pelleted.
13. Gently transfer the protoplast containing supernatant to another tube.
14. Centrifuge the suspension for 5 min at 5,000 × *g*.
15. Decant the supernatant and re-suspend the protoplasts in 10 mL of PTC buffer.
16. Centrifuge the suspension for 5 min at 5,000 × *g*.
17. Decant the supernatant and re-suspend the protoplasts in PTC buffer to an OD_{600} of ~1.0.
18. Divide the protoplast suspension in aliquots of 0.4–1.4 mL (0.2 mL needed for one transformation experiment) and put them into the freezer (–80 °C).

3.5. Protoplast Transformation

1. Take the protoplasts out of the freezer and thaw them quickly (see Note 21), without heating them too much.
2. Put 200 μL of the thawed protoplast suspension in an Eppendorf tube.
3. Add the DNA (or ligation mixture) to the protoplast suspension and mix gently by pipetting up and down.
4. *Immediately* add 500 μL of the 35% PEG6000 solution and mix by gently pipetting up and down.
5. Leave the mixture at room temperature for 5 min.
6. Plate the mixture on R2 plates (see Note 22) and incubate the plates at 27–30 °C for 16–20 h, allowing the protoplasts to regenerate.

7. Overlay the plate with 1 mL dH_2O containing the appropriate antibiotic (see Note 17).
8. Spread the antibiotic solution evenly (see Note 18).

4. Notes

1. Addition of $CaCl_2 \cdot 2H_2O$ and $CuSO_4$ to the medium before autoclaving will result in precipitation.
2. When using divalent cationic-sensitive antibiotics (e.g. kanamycin, apramycin), Ca^{2+}, and Cu^{2+} can inhibit antibiotic activity. It is, therefore, sometimes desirable to omit these components from the medium. This is possible but will result in retarded growth of the *S. lividans* colonies.
3. *S. lividans* and many other *Streptomyces* spp. show a pelleted growth. The Potter cell homogeniser (also known as tissue grinder), consisting of a glass tube and a Teflon piston, is used to homogenise the pellets in the pre-cultures.
4. TSB is used in this chapter as the standard rich medium. However, many other rich media are possible for the growth of *Streptomyces* spp. (see ref. 10).
5. The syringe containing non-absorbent cotton wool will be used as a filter to separate the spores from mycelium fragments.
6. *E. coli* S17-1 cells are methylation proficient and can be used for the conjugational transfer of DNA to *S. lividans*. When using other species, it might be necessary to use the methylation-deficient ET12567 strain.
7. Using cheap, regular soya flour from any supermarket yields the same result as using the expensive material from a laboratory product supplier.
8. For *S. lividans*, 0.8% of glycine is added to the culture. This concentration can vary from 0.5 to 1% for other *Streptomyces* spp.
9. While PEG6000 can be bought from different suppliers, it should be noted that testing PEG from different suppliers results in large differences in protoplast transformation efficiency.
10. Spore suspensions can also be inoculated directly into 5 mL phage medium (~10^6 spores), and the same holds true for those in glycerol stocks (50–100 μl). This is a more logical step if there is no later need for colonies on plate, since it allows researchers to skip growing *S. lividans* on plate which takes 2–3 days.

11. This culture can also be used for preliminary tests (e.g. enzymatic activity), but it is important to note that the phage medium is mostly geared towards biomass formation and that other media (such as TSB) are better suited for enzyme/secondary metabolite production.
12. In general, the pelleted growth of *S. lividans* does not interfere with later experiments. Should there be a problem after all, it is possible to reduce the pellet formation by using flasks containing baffles. A simple baffle can be obtained by inserting a stainless steel spring (30 cm length, 1.3 cm diameter, 19sw gauge) into an Erlenmeyer flask. If this still does not sufficiently solve the problem, addition of 34% sucrose, which *S. lividans* cannot catabolise, can further reduce pellet formation. Finally, it is also possible to achieve a more dispersed growth by use of other media, containing PEG8000 or Junlon, both of which favour dispersed growth (see ref. 10).
13. Incubate the plates until the entire plate has the dark-grey colour of *Streptomyces* spores. A white or light grey colour indicates the presence of non-sporulating aerial mycelium which will result in a lower final yield.
14. When starting from a glycerol stock, of spores inoculate into 3 mL LB-medium + antibiotic and grow for 5–6 h. Dilute this pre-culture for the overnight incubation described above.
15. It is advisable to start with step 8 in parallel once one put the samples in the centrifuge.
16. Drying the MS plates for 1 h in a laminar flow hood before plating greatly helps absorbing the 1 mL of solution added in the following step.
17. Apramycin and/or kanamycin resistance are the most often used selection markers; for these, add 1 mg per plate. For thiostrepton, add 750 μg per plate. When using apramycin, add the nalidixic acid first to the solution.
18. Spreading the solution can be done either with a spreader, very lightly, barely touching the plate or by manually, gently shaking of the plate.
19. The time needed for lysozyme treatment to form a sufficient amount of protoplasts varies greatly between strains. For *S. lividans*, we find that 15–20 min is usually enough, while other strains (e.g. *S. coelicolor*) may require up to 60 min. It should also be noted that *S. lividans* will not lyse when incubated for longer time periods (up to 1 h), but the same does not hold true for all *Streptomyces* spp.
20. Protoplasts will be visible under the light microscope as spheres among the mycelial clumps of un-protoplasted *S. lividans*. The protoplasts should fill the majority of the microscopic

field. If there are few protoplasts visible, a prolonged incubation is advised, since otherwise the final yield will generally be poor.

21. Thawing the protoplasts is best done either by placing the frozen tube in a warm water bath (40–45 °C) or by gently shaking the frozen tube under running warm water.

22. For ligation mixtures, we routinely plate one mixture on four R2 plates (4 × 150 μL). In the case of transformation with pure DNA, it might be desirable for the researcher to make tenfold dilutions in PTC and then plate the protoplasts.

References

1. Gabor, E.M., Alkema, W.B., and Janssen, D.B. (2004) Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ Microbiol* 6, 879–886.
2. Lorenz, P. and Eck, J. (2005) Metagenomics and industrial applications. *Nat Rev Microbiol* 3, 510–516.
3. Vrancken, K. and Anné, J. (2009) Secretory production of recombinant proteins by *Streptomyces*. *Future Microbiol* 4, 181–188.
4. Wang, G.Y., Graziani, E., Waters, B., Pan, W., Li, X., McDermott, J., et al. (2000) Novel natural products from soil DNA libraries in a streptomycete host. *Org Lett* 2, 2401–2404.
5. Courtois, S., Cappellano, C.M., Ball, M., Francou, F.X., Normand, P., Helynck, G., et al. (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl Environ Microbiol* 69, 49–55.
6. Sianidis, G., Pozidis, C., Becker, F., Vrancken, K., Sjoeholm, C., Karamanou, S., et al. (2006) Functional large-scale production of a novel *Jonesia* sp. xyloglucanase by heterologous secretion from *Streptomyces lividans*. *J Biotechnol* 121, 498–507.
7. Meilleur, C., Hupe, J.F., Juteau, P., and Shareck, F. (2009) Isolation and characterization of a new alkali-thermostable lipase cloned from a metagenomic library. *J Ind Microbiol Biotechnol* 36, 853–861.
8. MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H., and MacNeil, T. (1992) Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* 111, 61–68.
9. Martinez, A., Kolvek, S.J., Yip, C.L., Hopke, J., Brown, K.A., MacNeil, I.A., et al. (2004) Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Appl Environ Microbiol* 70, 2452–2463.
10. Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A. (2000) Practical *Streptomyces* Genetics. John Innes Center, Norwich, UK, pp. 1–613.

Chapter 7

Heterologous Gene Expression in the Hyperthermophilic Archaeon *Sulfolobus solfataricus*

Angel Angelov and Wolfgang Liebl

Abstract

One of the few available systems for gene expression in (hyper)thermophilic *Archaea* is the virus-based shuttle vector pMJ05 for *Sulfolobus solfataricus*. Although it is still not fully developed and there are some difficulties arising from the large size of the vector (>20 kb), it has successfully been used for the production of foreign and own proteins in *S. solfataricus*. Most often, the development of genetic tools for *Archaea* is held back by the lack of an efficient transformation system. In the case of pMJ05, this difficulty has been alleviated by using the *Sulfolobus* virus SSV1 as the vector backbone. The ability of the pMJ05 plasmid to spread in the culture as a virus, the availability of an effective selection marker (*pyr*) and of tunable promoters (*ara*S and *tf55*α) make this system one of the first choices for heterologous gene expression in (hyper)thermophilic *Archaea*.

Key words: Gene expression, Genetic system, Archaea, Hyperthermophilic, *Sulfolobus*

1. Introduction

Protein overproduction is a basic prerequisite for functional and protein–protein interaction studies. While a huge variety of gene expression systems exist for members of the eukaryal and bacterial kingdoms, the use of *Archaea* as gene expression hosts has been hampered by difficulties in their cultivation and the lack of efficient transformation tools. So far, only a few archaeal species are open to genetic manipulation (1). Most of them are mesophilic, for example, the halophilic archaeal genera *Haloferax* (2) and *Halobacterium* (3) and the methanogenic species *Methanococcus maripaludis* (4). So far, stable host–vector expression systems for hyperthermophilic archaea have been developed for representatives of the *Thermococcus* (5–7), *Pyrococcus* (8), and *Sulfolobus* (9–11) genera. Maybe the best developed and easily accessible of

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668,
DOI 10.1007/978-1-60761-823-2_7,

these is the virus-based shuttle vector series pMJ for *S. solfataricus* (10). These vectors have been developed on the basis of the virus SSV1, initially isolated as a 15 kb plasmid from *Sulfolobus shibatae* (12) and later shown to be able to infect also *S. solfataricus* (13). Wild-type viral DNA is packed in lemon-shaped particles, and upon infection, the DNA was found to integrate site-specifically in the host chromosome and in addition was detected to replicate as a plasmid with three to five copies per chromosome. The pMJ vectors have been constructed by combining the whole SSV1 viral DNA with the pUC18 plasmid for propagation in *E. coli* and the *S. solfataricus pyrEF* genes for selection of transformants in pyrimidine auxotrophic *S. solfataricus* strains. Initial introduction of these vectors in *S. solfataricus* proceeds via electroporation, but subsequently the vectors propagate throughout the culture as a virus, thus eliminating the need for an efficient transformation procedure. High-level gene expression is achieved by using one of two different promoters, the heat-inducible *tf55α* promoter from the major chaperonin, thermophilic factor 55, and the arabinose-inducible *araS* promoter from the arabinose-binding protein, AraS (14). The system has already successfully been applied for the production of recombinant and His- or Strep-tagged proteins of different origin in *S. solfataricus* (9).

2. Materials

2.1. Cloning Procedures

Several strategies for cloning of the Gene of Interest (GOI) in pMJ05 are possible. Here, one option is used, provided that the GOI does not contain *Nco*I, *Apa*I, *Bln*I, and *Eag*I restriction sites (Fig. 1):

1. Entry vector pSVA5 (10).
2. Shuttle vector pMJ05 (10).
3. Primers to amplify the GOI, introducing *Nco*I site at the 5′-terminus and *Apa*I site at the 3′-terminus.
4. The following buffer is suitable for performing the restriction reaction with *Nco*I and *Apa*I simultaneously: 10 mM Tris–HCl (pH 7.5 at 37°C), 10 mM $MgCl_2$, 0.02% Triton X-100, 0.1 mg/mL BSA.
5. Enzymes and reagents commonly used for performing molecular cloning, e.g., proofreading DNA polymerase, restriction endonucleases, T4 ligase, etc.

2.2. Strains and Growth Media

1. *S. solfataricus* PH1–16 (15).
2. *E. coli* DH10B (16).

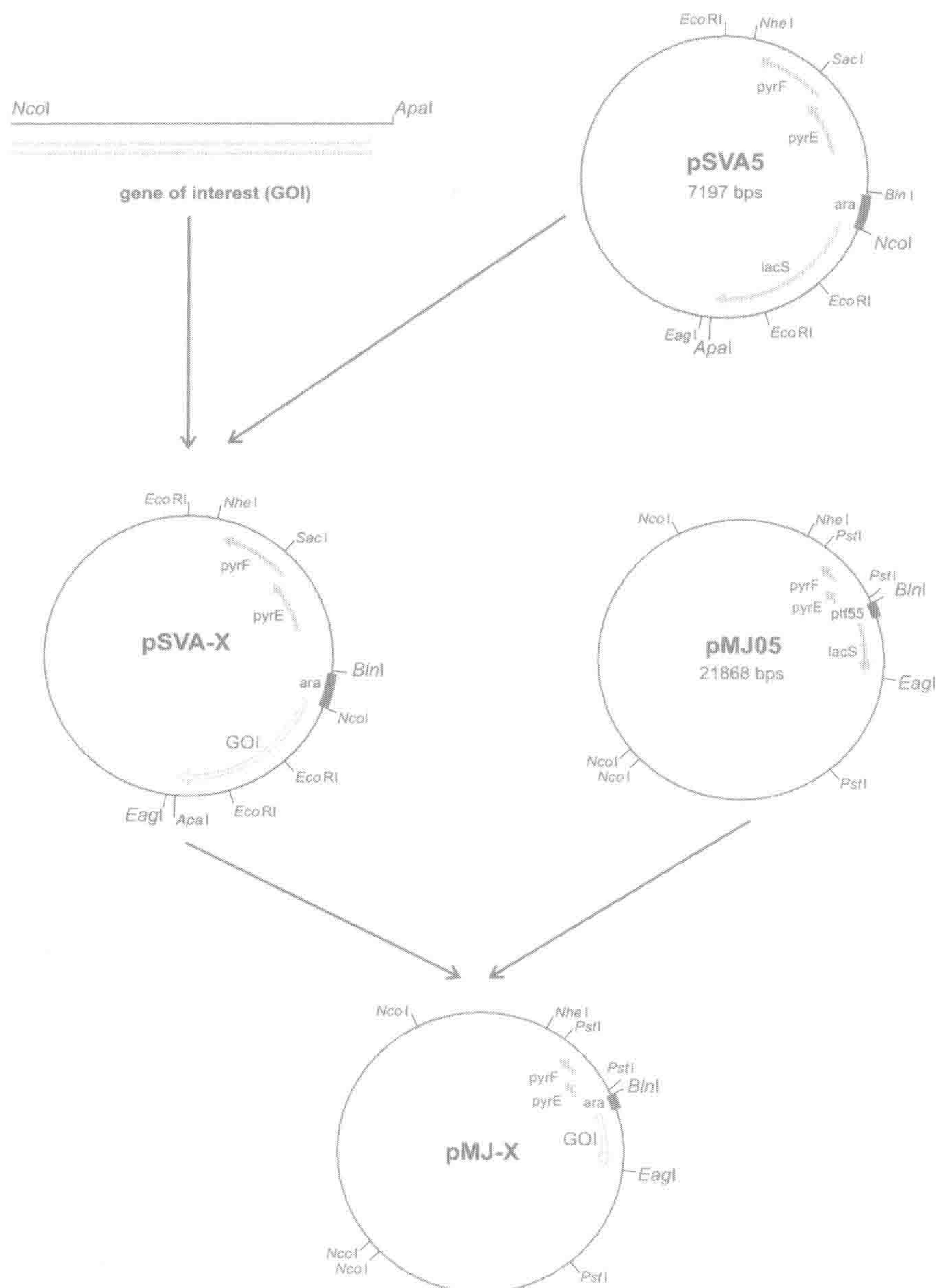


Fig. 1. A scheme of the steps used for cloning of a PCR amplified GOI in the pMJ05 vector. The restriction sites relevant to the cloning steps are in bold; sites useful for analytical restriction reactions are also shown.

3. Brock's basal salts solutions (for 1 L):
   (a) 100× Brock's: 130 g $(NH_4)_2SO_4$, 25 g $MgSO_4 \cdot 7H_2O$, 2 g $FeCl_3 \cdot 6H_2O$, 3 mL 50% H_2SO_4, sterilize by filtration.
   (b) 200× Brock's: 56 g KH_2PO_4 and the following volumes of 10 mg/mL stock solutions of the trace elements:

36 mL $MnCl_2$, 4.4 mL $ZnSO_4$, 1 mL $CuCl_2$, 0.6 mL $VOSO_4$, 0.2 mL $CoSO_4$, 90 μL $Na_2B_4O_7$, 6 μL Na_2MoO_4; autoclave and add 5 mL 50% H_2SO_4.

(c) 1,000× Brock's: 70 g $CaCl_2 \cdot 2H_2O$; autoclave.

4. *S. solfataricus* growth medium: Add 700 mL H_2O_{bidest} to a 1 L flask and autoclave. After cooling, add 10 mL 100×, 5 mL 200×, and 1 mL 1,000× Brock's salt solutions. Add 5 mL of 20% tryptone solution and bring to 1 L with sterile water. Adjust pH to 3–3.5 with 50% H_2SO_4.
5. *E. coli* growth media:
   (a) LB: 10 g tryptone, 5 g yeast extract, and 5 g NaCl; bring to 1 L with H_2O and sterilize by autoclaving.
   (b) SOC: 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, and 2.5 mL 1 M KCl; bring to 1 L with H_2O and sterilize by autoclaving.
   (c) For the preparation of plates, add 1.5% agar before autoclaving. When needed, supplement with ampicillin at 100 μg/mL final concentration.
6. Water-bath shaker for growing cultures of *S. solfataricus* filled with Rotitherm K+H (Carl Roth GmbH, Karlsruhe) and 100 mL Erlenmeyer flasks with long necks (see Note 1).

2.3. Transformation of S. solfataricus

1. Electroporation equipment, for example, Gene Pulser (Bio-Rad). The electroporation cuvettes should have an electrode distance of 0.1 cm.
2. Approximately 100 mL of 20 mM saccharose solution.
3. Heat block with a mixing function (thermomixer).

2.4. Analysis of Transformants by SDS-PAGE

1. 10 mM Tris–HCl pH 8.0 for the preparation of crude cellular extract of *S. solfataricus*.
2. Separating buffer (4×): 1.5 M Tris–HCl pH 8.7, 0.4% SDS. Store at room temperature.
3. Stacking buffer (4×): 0.5 M Tris–HCl pH 6.8, 0.4% SDS. Store at room temperature.
4. 30% acrylamide/bisacrylamide solution (37.5:1) and *N,N,N,N'*-Tetramethyl-ethylenediamine (TEMED).
5. Ammonium persulfate: prepare 10% solution in water and freeze in single-use aliquots at −20°C.
6. Running buffer (5×): 125 mM Tris, 960 mM glycine, 0.5% (w/v) SDS. Store at room temperature.
7. Prestained molecular-weight marker.
8. Gel staining solution: Coomassie Blue R-250 0.05% (w/v), acetic acid 10% (v/v), isopropanol 25% (v/v).

3. Methods

3.1. Cloning of the Gene of Interest in the Vector pMJ05 (with an araS Promoter)

Various cloning strategies can be used to introduce the GOI in the shuttle vector pMJ05. Most easily, this can be achieved by using the following strategy (Fig. 1):

1. Amplify the GOI using a proofreading DNA polymerase (e.g., Pfu) and primers that introduce *Nco*I site at the 5′-terminus and *Apa*I site at the 3′-terminus.
2. Perform restriction digest of the PCR product with *Nco*I and *Apa*I.
3. Ligate the restricted PCR fragment with the *Nco*I and *Apa*I digested and dephosphorylated entry vector pSVA5. In this way, the *lacS* gene of pSVA5 is replaced with the GOI.
4. Perform restriction digest of the newly derived pSVA-X vector with *Bln*I and *Eag*I, excise the resulting fragment from agarose gel and purify it.
5. Ligate the purified *Bln*I–*Eag*I DNA fragment with the pMJ05 vector, which has been digested with the same enzymes and dephosphorylated (see Note 2).

An important issue when working with pMJ05 and derivatives is that the cloning steps and vector maintenance should be performed in *E. coli* DH10B, and the plasmid carrying cells should be incubated at 28°C to avoid recombination events.

3.2. Transformation of S. solfataricus

The final construct is introduced in *S. solfataricus* via electroporation using parameters that have been optimized specifically for this organism. It is advisable to always include appropriate controls in the transformation reactions, e.g., a no DNA control and pMJ05 without the GOI (see Note 3):

1. The cells of a 50 mL overnight culture of *S. solfataricus* PH1–16 (OD_{600nm} ~ 0.1–0.3) are cooled on ice for 15 min and then collected by centrifugation at 4,000 × *g* at 4°C. Medium salts are removed by successive washing of the cells in 50, 25, and 1 mL ice-chilled 20 mM saccharose solution. Following the final wash step, the cells are resuspended in saccharose solution at a concentration of 10^{10} cells/mL and are kept on ice until use.
2. Aliquots of 50 μl cells are added to the precooled electroporation cuvettes and are carefully mixed with 1 μl of plasmid DNA (maximally 300 ng). Electroporation is performed with the following parameters:

   Voltage: 1.5 kV

   Capacity: 25 μF

   Resistance: 400 Ω

The highest transformation efficiency is achieved at a time constant of 9.1 ms. One milliliter of growth medium is added immediately after the electroshock, the cells are transferred to 1.5 mL reagent tubes and regenerated for 2 h at 75°C with moderate shaking. Finally, the transformation mixture is transferred to 50 mL growth medium in 100 mL Erlenmeyer flasks and is further incubated for 2 days at 75°C in a water-bath shaker.

3. The expression of the GOI is induced by transferring of 10 mL *S. solfataricus* cells from the primary transformation mixture to 100 mL of growth medium supplemented with 0.2% arabinose. After 2 days of growth, the cells are harvested by centrifugation and kept frozen.

3.3. Analysis of *S. solfataricus* Transformants by SDS-PAGE

S. solfataricus transformants can be analyzed by different means, depending on the nature of the recombinant protein. The crude cellular extract of the transformed *S. solfataricus* cells can be analyzed by SDS-PAGE before and after induction with arabinose. In the case when an enzyme is produced, specific enzyme activity of the transformants can be compared with the one in negative controls. Additionally, Southern blot analysis can be performed to confirm the chromosomal integration of the pMJ vector constructs:

1. The *S. solfataricus* cell pellet is resuspended in 10 mM Tris–HCl pH 8.0 (1 mL of buffer for the cell pellet of approximately 30 mL *S. solfataricus* culture).
2. One milliliter of the cell suspension is subjected to 5 min of sonification on ice-water bath.
3. The resulting crude cellular extract is centrifuged (10 min, 10,000 × *g*) and the supernatant (soluble fraction) is separated from the pellet (insoluble fraction).
4. The protein concentration in the samples is determined (for example, using Bradford reagent).
5. SDS-PAGE gels (12%) are prepared by mixing 5 mL of 4× separating buffer, with 8 mL acrylamide/bis solution, 12 mL water, 100 μL ammonium persulfate solution, and 20 μL TEMED. After pouring, the gels are overlaid with isopropanol and left to polymerize for about 30 min. After pouring off the isopropanol, the stacking gel is applied (prepared by mixing 2.5 mL of 4× stacking buffer with 1.3 mL acrylamide/bis solution, 6.1 mL water, 50 μL ammonium persulfate solution, and 10 μL TEMED).
6. The protein samples are applied (15–20 μg protein/lane) after mixing with loading dye and boiling for 5 min. The gels are run at 20–30 mA until the dye front reaches the end of the glass plates and are stained overnight in gel staining solution. The gels are then destained several times in 10% acetic acid.

4. Notes

1. The Erlenmeyer flasks with long necks are not commercially available, but ordinary ones can be modified by any glass manufacturer. The long necks (about 40 cm) remain above the water bath during the incubation at 75°C and thus enable condensation and backflow of evaporated culture medium.
2. Manipulations of the pMJ05 vector are to be performed carefully with regard to its large size, e.g., 21.8 kb. For example, DNA purification from an agarose gel using commercially available ion exchange columns has to be performed with keeping in mind that most of these have an upper cutoff limit of approximately 20 kb. Therefore, isolation from Low Melting Point (LMP) agarose is recommended with the help of a β-agarose digesting enzyme (for example, GELase, Epicentre).
3. The two control reactions suggested here should be treated in the same manner as the normal transformation reactions, in the first reaction omitting DNA and in the second reaction adding the shuttle vector pMJ05. The "no DNA" control transformation mixture is then split and inoculated in two different *Sulfolobus* media: one with and one without the addition of uracil at 5 μg/mL. The uracil-supplemented sample should show normal growth, and the one lacking uracil should show no or very weak growth after incubation for 2 days at 75°C (the *Sulfolobus* strain PH1–16 is auxotrophic for uracil). The positive control reaction (pMJ05) is expected to show normal growth without uracil supplementation.

Acknowledgements

The authors wish to thank Christa Schleper and Melanie Jonuscheit for the kind introduction to the pMJ05 system.

References

1. Allers, T. and Mevarech, M. (2005) Archaeal genetics– the third way. *Nat Rev Genet* 6, 58–73.
2. Kaczowka, S.J., Reuter, C.J., Talarico, L.A., and Maupin-Furlow, J.A. (2005) Recombinant production of *Zymomonas mobilis* pyruvate decarboxylase in the haloarchaeon *Haloferax volcanii*. *Archaea* 1, 327–334.
3. Heymann, J.A., Havelka, W.A., and Oesterhelt, D. (1993) Homologous overexpression of a light-driven anion pump in an archaebacterium. *Mol Microbiol* 7, 623–630.
4. Gardner, W.L. and Whitman, W.B. (1999) Expression vectors for *Methanococcus maripaludis*: overexpression of acetohydroxyacid synthase and beta-galactosidase. *Genetics* 152, 1439–1447.
5. Sato, T., Fukui, T., Atomi, H., and Imanaka, T. (2005) Improved and versatile transformation system allowing multiple genetic manipulations of the hyperthermophilic archaeon *Thermococcus kodakaraensis*. *Appl Environ Microbiol* 71, 3889–3899.

6. Sato, T., Fukui, T., Atomi, H., and Imanaka, T. (2003) Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J Bacteriol* 185, 210–220.
7. Santangelo, T.J., Cubonova, L., and Reeve, J.N. (2008) Shuttle vector expression in *Thermococcus kodakaraensis*: contributions of cis elements to protein synthesis in a hyperthermophilic archaeon. *Appl Environ Microbiol* 74, 3099–3104.
8. Lucas, S., Toffin, L., Zivanovic, Y., Charlier, D., Moussard, H., Forterre, P., et al. (2002) Construction of a shuttle vector for, and spheroplast transformation of, the hyperthermophilic archaeon *Pyrococcus abyssi*. *Appl Environ Microbiol* 68, 5528–5536.
9. Albers, S.V., Jonuscheit, M., Dinkelaker, S., Urich, T., Kletzin, A., Tampe, R., et al. (2006) Production of recombinant and tagged proteins in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Appl Environ Microbiol* 72, 102–111.
10. Jonuscheit, M., Martusewitsch, E., Stedman, K.M., and Schleper, C. (2003) A reporter gene system for the hyperthermophilic archaeon *Sulfolobus solfataricus* based on a selectable and integrative shuttle vector. *Mol Microbiol* 48, 1241–1252.
11. Aucelli, T., Contursi, P., Girfoglio, M., Rossi, M., and Cannio, R. (2006) A spreadable, non-integrative and high copy number shuttle vector for *Sulfolobus solfataricus* based on the genetic element pSSVx from *Sulfolobus islandicus*. *Nucleic Acids Res* 34, e114.
12. Palm, P., Schleper, C., Grampp, B., Yeats, S., McWilliam, P., Reiter, W.D., et al. (1991) Complete nucleotide sequence of the virus SSV1 of the archaebacterium *Sulfolobus shibatae*. *Virology* 185, 242–250.
13. Stedman, K.M., Schleper, C., Rumpf, E., and Zillig, W. (1999) Genetic requirements for the function of the archaeal virus SSV1 in *Sulfolobus solfataricus*: construction and testing of viral shuttle vectors. *Genetics* 152, 1397–1405.
14. Lubelska, J.M., Jonuscheit, M., Schleper, C., Albers, S.V., and Driessen, A.J. (2006) Regulation of expression of the arabinose and glucose transporter genes in the thermophilic archaeon *Sulfolobus solfataricus*. *Extremophiles* 10, 383–391.
15. Schleper, C., Kubo, K., and Zillig, W. (1992) The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: demonstration of infectivity and of transfection with viral DNA. *Proc Natl Acad Sci USA* 89, 7645–7649.
16. Grant, S.G., Jessee, J., Bloom, F.R., and Hanahan, D. (1990) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc Natl Acad Sci USA* 87, 4645–4649.

Chapter 8

Novel Tools for the Functional Expression of Metagenomic DNA

Sonja Christina Troeschel, Thomas Drepper, Christian Leggewie, Wolfgang R. Streit, and Karl-Erich Jaeger

Abstract

The functional expression of environmental genes in a particular host bacterium is hampered by various limitations including inefficient transcription of target genes as well as improper assembly of the corresponding enzymes. Therefore, the identification of novel enzymes from metagenomic libraries by activity-based screening requires efficient expression and screening systems. In the following chapter, we present two novel tools to improve the functional expression of metagenomic genes. (1) Comparative screenings of metagenomic libraries demonstrated that different enzymes were detected when phylogenetically distinct expression host strains were used. Thus, we have developed a strategy, which comprises library construction using a shuttle vector that allows comparative expression and screening of metagenomic DNA in *Escherichia coli*, *Pseudomonas putida*, and *Bacillus subtilis*. (2) Expression studies have revealed that functional expression of environmental genes in heterologous expression hosts is often limited by insufficient promoter recognition. Therefore, a method is described allowing to enhance the expression capacity of *E. coli* by using the transposon MuExpress. This recombinant transposon is able to insert randomly into environmental DNA fragments thereby facilitating gene expression from its two inducible promoters.

Key words: Metagenomic library, Environmental DNA, Activity-based screening, Functional expression, Multi-host screening, Shuttle vector, *Escherichia coli*, *Pseudomonas putida*, *Bacillus subtilis*, Transposon

1. Introduction

The discovery, cloning, and functional characterization of novel genes or gene clusters from uncultured microorganisms comprises metagenomics, which is defined as culture-independent genomic analysis of a microbial community (1). Sampling microbes or DNA from a given environment constitutes the first step of all metagenomic approaches. Different considerations influence the selection of a habitat to be sampled. In general, it may be reasonable to

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668,
DOI 10.1007/978-1-60761-823-2_8,

select either environments with a high microbial diversity such as soils or sediments (1, 2) or extreme environments that typically harbor few but highly specialized organisms perfectly adapted to the hostile conditions of the respective habitat (3). Therefore, the selection of the habitat predefines the diversity as well as general properties of putative biocatalysts to be discovered. For instance, to identify novel enzymes with specific properties (e.g., stability against high temperature, pH, pressure, or salt tolerance) extreme environments may be mined. However, in this context, one has to consider that intracellular enzymes of extremophiles are not necessarily adapted to these conditions, too (3, 4). In addition, it is also possible to select an environment that is naturally enriched for target biocatalysts (5) or to enrich samples in the laboratory, e.g., by iterative incubation cycles in the presence of defined substrates, which are expected to select for the enzymatic activity of choice (4, 6–8). Both strategies increase the chance to find genes of interest. Nevertheless, a major drawback of employing any enrichment steps is the loss of microbial diversity by favoring fast growing and culturable parts of microbial consortia (9–11).

Basically, two fundamentally different approaches, namely the sequence-dependent and -independent screening, can be applied to detect novel enzymes in metagenomic libraries. Both approaches have their limitations: sequence-dependent screenings rely on DNA- or protein-sequence comparisons and usually allow the identification of novel enzymes which share homology to already known ones (e.g., reviewed in 9, 12, 13), whereas sequence-independent approaches require efficient expression and screening systems (14, 15). In this chapter, we will focus on the construction of metagenomic libraries and their activity-based screening.

Most activity screenings of metagenomic libraries are based on the cultivation of metagenomic clones on indicator plates allowing analysis of defined enzyme activities via biocatalytic conversion of an indicator substrate that leads to the formation of a clear or colored halo surrounding the "positive" colony. However, the frequency of active clones is usually quite low and strongly depends on the sensitivity of the used assay system as well as on the gene expression capability (10, 14–18). Therefore, successful construction and screening of metagenomic libraries are affected by various parameters including (1) the sample composition, collection, and storage, (2) an unbiased nucleic acid extraction, purification, and fragmentation, (3) the relationship between DNA fragment size and gene cluster length, and finally, (4) the gene distribution and library size (9, 13, 19).

In principle, activity-based screening of metagenomes basically requires the concerted expression of all environmental genes located on a given DNA fragment, irrespective of its size and structure. Thus, recognition of promoters and/or regulatory elements by the transcription machinery of the expression host is essential for

the functional expression of the library genes (15, 20, 21). Alternatively, host-specific promoters can be placed upstream and/or downstream of the environmental DNA within a library vector to functionally express the heterologous genes (16). However, the position or orientation of a target gene relative to the host promoter as well as the over-all insert size may dramatically limit its expression efficiency. In addition, many factors such as chaperones, cofactors, posttranslational modifications, and secretion systems might be needed for correct assembly and targeting of the recombinant polypeptides (22). Thus, novel techniques and tools have been established to further improve the expression of genes from metagenomic libraries.

Comparative screenings of metagenomic libraries have shown that different types of active enzymes were detected when phylogenetically distinct hosts were used. This "different host–different hit" effect (19) may be attributed to differences in transcription and translation efficiencies, as well as to individual metabolic properties of the used host organisms. Hence, broad-host-range vectors allowing efficient cloning of metagenomic DNA in common "cloning hosts" as *Escherichia coli*, and subsequent transfer and maintenance of the metagenomic library in different "screening hosts" greatly simplify multiple library construction and screening and thus significantly enhance detection frequencies for genes that confer novel functions (23–25). Furthermore, transcription of metagenomic genes can be initiated or enhanced even after construction of a metagenomic library. To this end, transposon MuExpress was constructed, which randomly integrates into environmental DNA libraries and permits the T7 RNA polymerase-dependent expression of its flanking genes (26). Insertion of MuExpress into a metagenomic library derived from a drinking water biofilm (27) resulted in identification of a previously not identified clone exhibiting lipolytic activity.

In the following section, we illustrate a method to isolate and purify metagenomic DNA from bacterial consortia. It has to be mentioned that the protocol should not be used to extract DNA from soil samples, because humic substances, organic compounds, or saline that often contaminate environmental DNA are difficult to remove and interfere with enzyme reactions like restriction digest or ligation. Subsequently, we will describe the fragmentation of DNA by restriction endonucleases, the preparation of the expression vector, and the construction of a metagenomic library. This procedure is schematically visualized in Fig. 1. Function-based screenings of libraries always require expression of the target gene. Therefore, two different strategies will be presented allowing to improve heterologous gene expression, namely comparative expression of metagenomic genes in *E. coli*, *Pseudomonas putida*, and *Bacillus subtilis* by using the newly constructed shuttle vector pEBP18 (S.C. Troeschel, unpublished data) and MuExpress-mediated expression of environmental genes in *E. coli*.

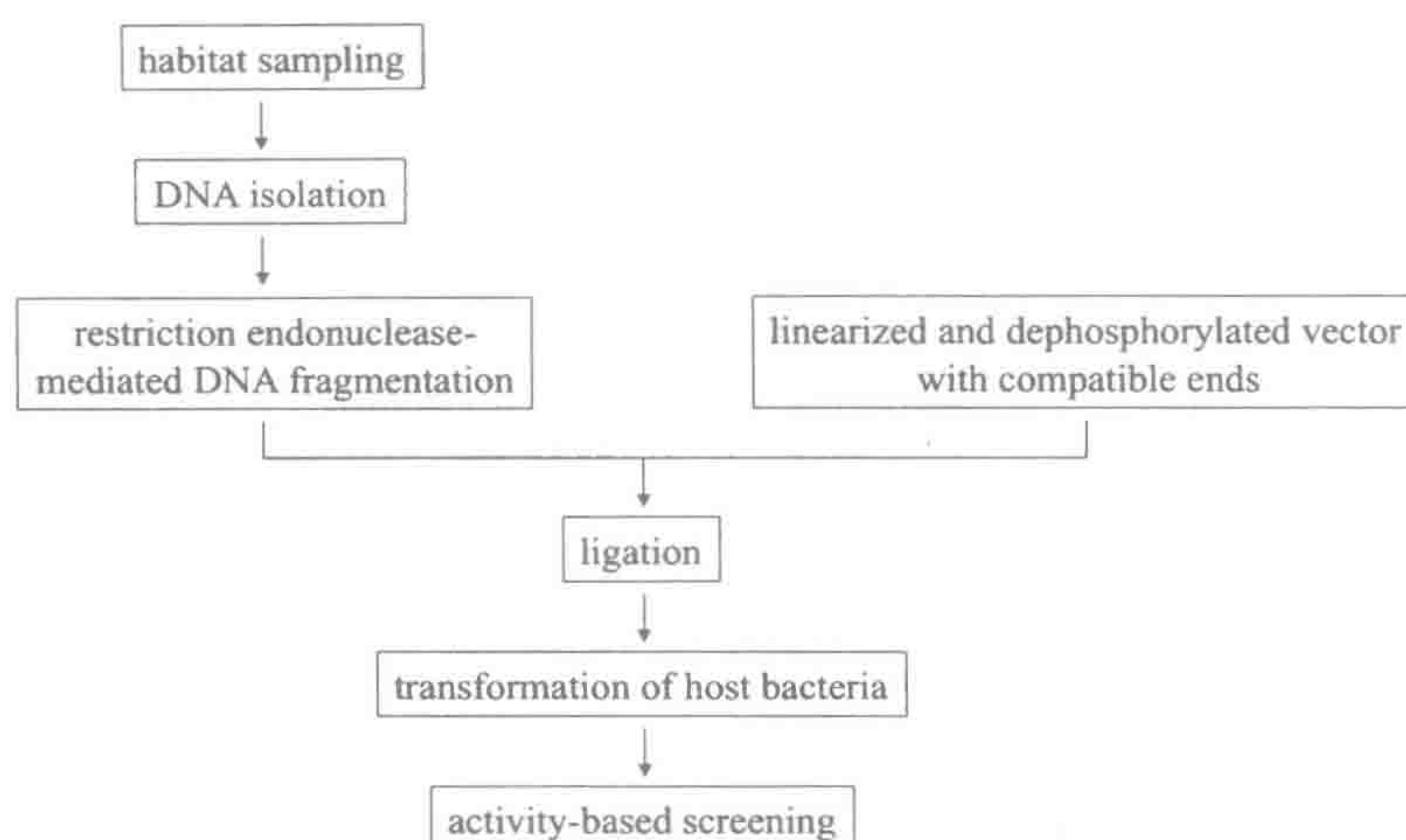


Fig. 1. Schematic outline of the experimental steps to construct a metagenomic library.

2. Materials

2.1. Bacterial Strains, Media, and Antibiotics

Chemicals were obtained from Carl Roth GmbH (Karlsruhe, Germany) if not noted otherwise.

1. *E. coli* DH5α or DH10B (Invitrogen GmbH, Karlsruhe, Germany) and *E. coli* BL21(DE3) (Novagen, Merck KGaA, Darmstadt, Germany) are used for DNA cloning and (T7 RNA polymerase-dependent) expression.
2. *P. putida* KT2440 (28) and *B. subtilis* TEB1030 (29) are used for screening.
3. Luria-Bertani (LB) medium consisting of 10 g/L trypton (peptone from casein), 5 g/L yeast extract, and 5 g/L NaCl is prepared as described by Sambrook and Russell (30).
4. EM medium: 20 g/L trypton (peptone from casein), 5 g/L yeast extract, and 5 g/L NaCl are solubilized in deionized water and pH is adjusted to 7.2. After sterilization (121°C, 2 bar, 20 min), 5 mL/L sterile glucose solution (50% (w/v) α-D(+)-glucose monohydrate) is added to obtain a final concentration of 0.5% (w/v).
5. EM1 is prepared analogous to EM, but contains 1% (w/v) glucose.
6. If necessary, antibiotics are added to culture media to a final concentration of 150 μg/mL ampicillin (ampicillin sodium salt), 20 μg/mL kanamycin (kanamycin sulfate), or 7.5 μg/mL chloramphenicol.
7. Agar plates: liquid medium supplemented with 1.5% (w/v) agar prior to sterilization.

8. Starch plates: LB agar supplemented with 1% (w/v) corn starch (Mondamin, discounter) before sterilization. Starch plates are colored with iodine solution: 0.5% (w/v) iodine (Fluka, Sigma-Aldrich Chemie GmbH, Munich, Germany) and 1% (w/v) potassium iodine (AppliChem GmbH, Darmstadt, Germany) solubilized in deionized water (solubilization needs several days!).

2.2. Vectors

1. pUC18: 20 ng/μL, carrying an ampicillin resistance gene (Fermentas GmbH, St. Leon-Rot, Germany).
2. pBBR1MCS2: 20 ng/μL, carrying a kanamycin resistance gene (31).
3. Shuttle vector pEBP18 (Fig. 2). pEBP18 replicates in *E. coli* (ori_{Ec}) and in *P. putida* (ori_{Pp}) as episomal plasmid. In *B. subtilis*, the shuttle vector integrates via homologous recombination into the amylase locus ($amyE'_{Bs}$, $'amyE_{Bs}$). Heterologous DNA is introduced into the *Bam*HI cloning site and can be reisolated by *Swa*I (*Smi*I) digestion. Heterologous genes can be expressed using the inducible promoters P_{T7} (T7 RNA polymerase-dependent) or P_{xyl} (xylose). Gene transcription can be monitored by GFP (*gfp*) expression. The shuttle vector also possesses a *cos* site allowing highly efficient transduction of

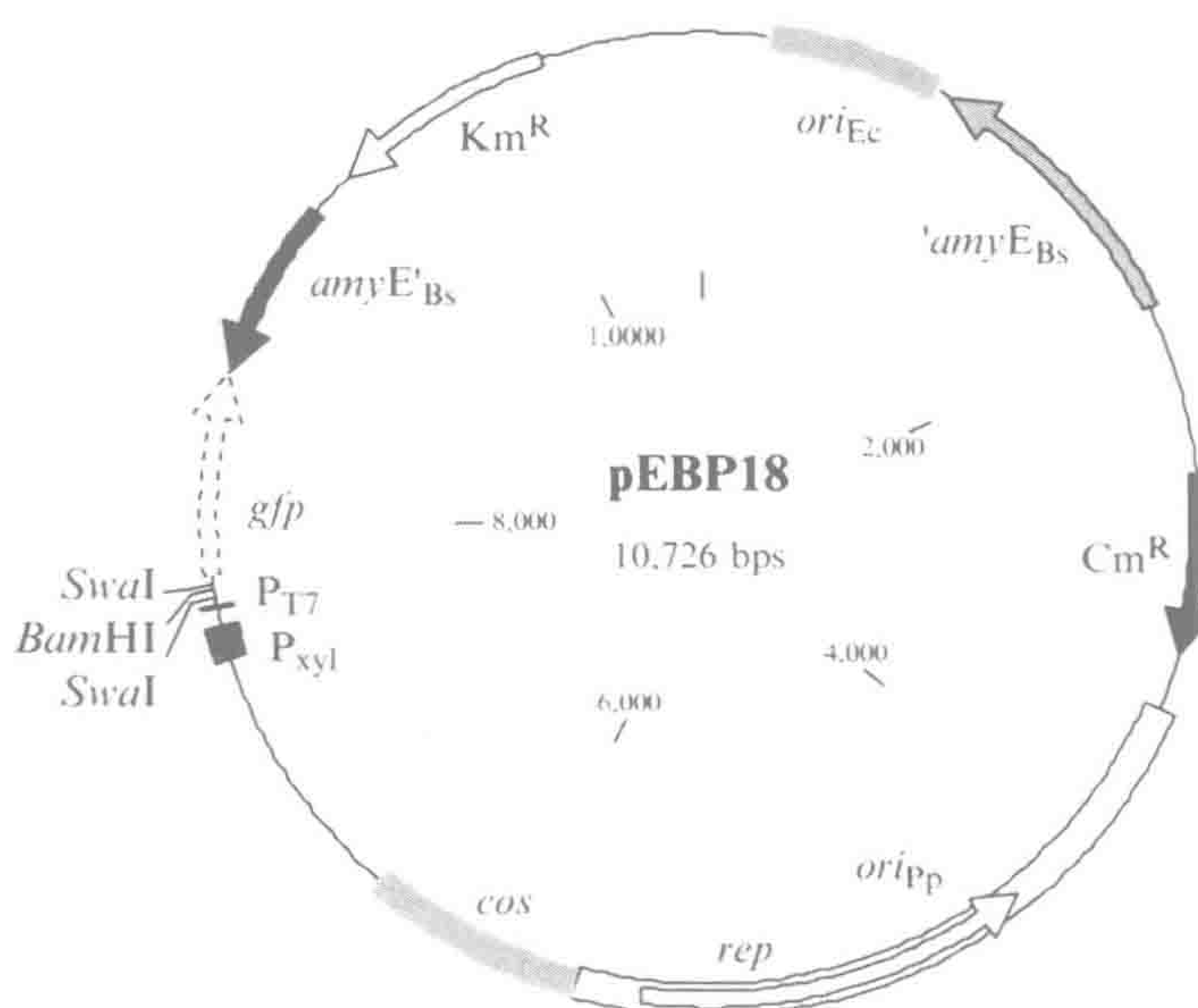


Fig. 2. Shuttle vector pEBP18 replicates in the host organisms *Escherichia coli*, *Pseudomonas putida*, and *Bacillus subtilis*. Heterologous DNA is cloned into the singular *Bam*HI site and can be reisolated by *Swa*I digestion. The color code assigns the different vector features to the respective host organism. *Black*: *B. subtilis*, *dark grey*: *E. coli*, *white*: *P. putida*, *light grey*: *E. coli*, and *P. putida dashed line*: *GFP*; for more detailed information see text.

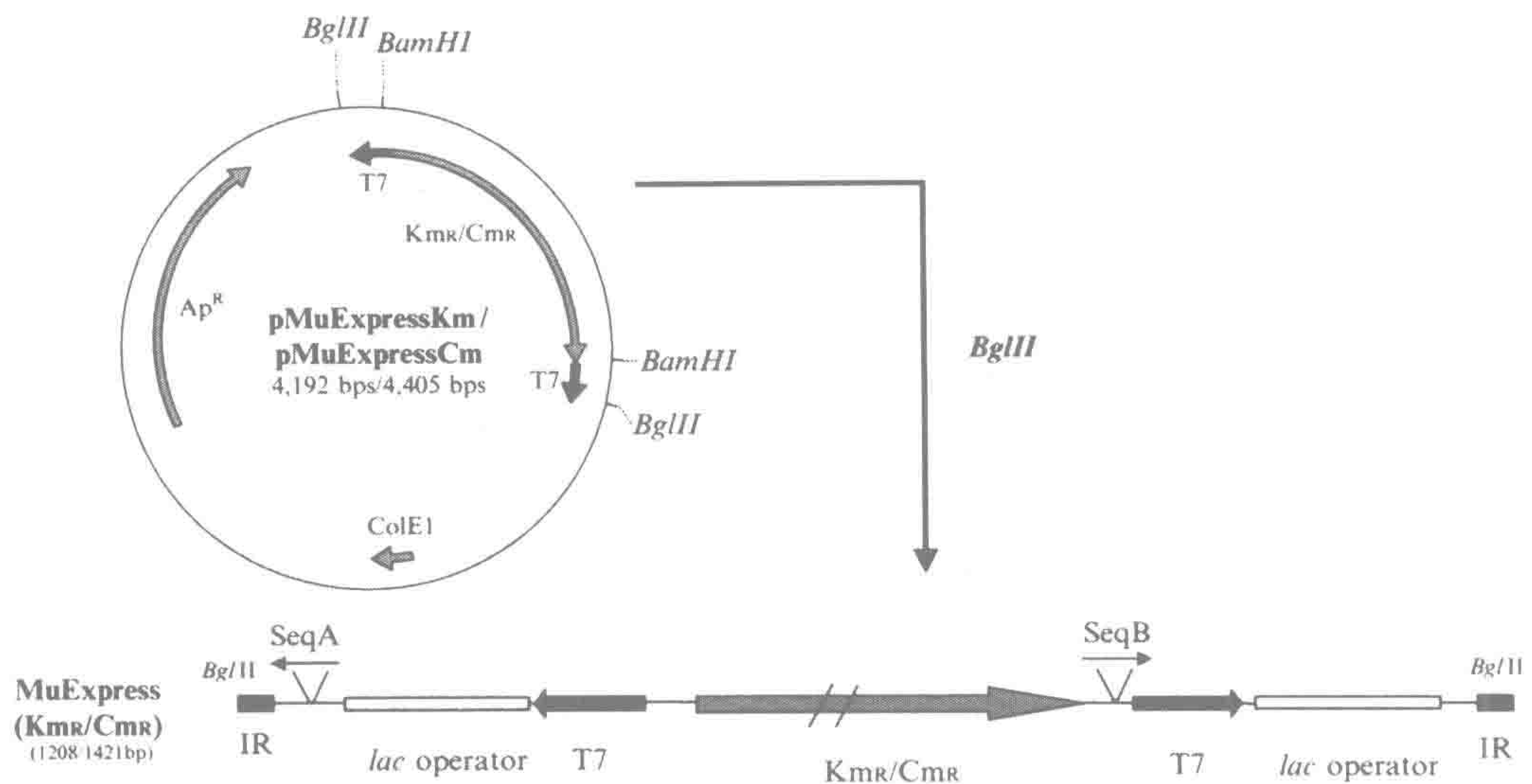


Fig. 3. Transposon MuExpress can be isolated by digestion of plasmid pMuExpress with *Bgl*II. *Km* kanamycin resistance gene, *Cm* chloramphenicol resistance gene, *Ap* ampicillin resistance gene, *T7* T7-promoter region, *IR* binding sites for transposase, *SeqA/B* sequencing primer sites, drawing is not to scale.

E. coli by phage infection, which is used if the shuttle vector carries large inserts increasing its size to 37–52 kb (32).

4. pMuExpress (Fig. 3). The vectors pMuExpressKm and pMuExpressCm contain the transposons MuExpressKm and MuExpressCm, respectively. The vector backbone confers ampicillin resistance and replication is initiated by a ColE1 origin. The transposon can be excised from vector pMuExpress by using two flanking *Bgl*II sites. MuExpress harbors two strong T7 promoters for gene expression. Additionally, the recombinant transposon allows the bidirectional sequencing of active clones starting from unique primer binding sites SeqA or SeqB.

2.3. Preparation of Competent Cells: Solutions and Cuvettes

1. Chemical competent *E. coli* DH5α cells are available from Invitrogen GmbH (Karlsruhe, Germany). Alternatively, prepare chemical competent *E. coli* according to ref. 30, Vol. 1, Chap. 1 including the calcium chloride method (protocol 25), Hanahan methods (protocol 23), or Inoue method (protocol 24).
2. My-budget electroporation cuvettes: 1 mm gap and 2 mm gap (Bio-Budget Technologies GmbH, Krefeld, Germany).
3. MilliQ water is a registered trademark for water purification systems manufactured by Millipore GmbH (Schwalbach, Germany) providing water with a conductance of 18.2 mΩ at 25°C. After sterilization (121°C, 2 bar, 20 min) store at 4°C.

4. Glycerol solution: 10% (v/v) Rotipuran glycerol (Carl Roth GmbH, Karlsruhe, Germany) in MilliQ water. After sterilization (121°C, 2 bar, 20 min) store at 4°C.
5. Sucrose solution: 300 mM sucrose (Merck KGaA, Darmstadt, Germany) in MilliQ water. After sterilization (121°C, 2 bar, 20 min) store at room temperature.
6. Paris-Medium: 60 mM potassium hydrogen phosphate (K_2HPO_4), 40 mM potassium dihydrogen phosphate (KH_2PO_4), 3 mM trisodium citrate dihydrate (Na_3-citrat), 20 mM potassium L-glutamate monohydrate (K-L-glutamate, AppliChem GmbH, Darmstadt, Germany), 3 mM magnesium sulfate ($MgSO_4$), 1% (w/v) α-D(+)-glucose, 0.1% (w/v) Bacto casamino acids (Becton Dickinson GmbH, Heidelberg, Germany), 20 mg/L L-tryptophan (Merck KGaA, Darmstadt, Germany), 2.2 mg/L ammonium-iron (III) citrate (Fe(III) NH_4-citrate, Fluka, Sigma-Aldrich Chemie GmbH, Munich, Germany). Paris-Medium is prepared from separately sterilized stock solutions according to Table 1 and can be stored for 2 weeks at 4°C. All chemicals are obtained from Carl Roth GmbH (Karlsruhe, Germany) if not noted otherwise.

Table 1
Paris-Medium prepared from stock solutions

Solution	[Stock][a]	Sterilization[b]	Storage	Volume
K_2HPO_4	0.5 M	121°C	RT	6 mL
KH_2PO_4	1 M	121°C	RT	2 mL
Na_3-citrat	0.5 M	121°C	RT	300 µL
K-L-glutamate	1 M	121°C	RT	1 mL
$MgSO_4$	1 M	121°C	RT	150 µL
Glucose	50% (w/v)	121°C	RT	1 mL
Casamino acids	10% (w/v)	0.22 µm	-20°C	1 mL
L-Tryptophan	5 mg/mL	0.22 µm	-20°C	200 µL
Fe(III)NH_4-citrate	2.2 g/L	0.22 µm	+4°C	2.5 mL
MilliQ water		121°C	RT	Add 50 mL

RT room temperature
[a]The concentration of stock solution is given by [stock]
[b]121°C, autoclave at 121°C; 2 bar, 20 min; 0.22 µm, filtrate with a sterile filter having a pore diameter of 0.22 µm

2.4. Solutions for DNA Extraction and Purification

Chemicals were obtained from Carl Roth GmbH (Karlsruhe, Germany) if not noted otherwise.

1. Isopropanol: Rotisolv 2-propanol.
2. Ethanol: Rotisolv ethanol is diluted to 70% (v/v) using deionized water.
3. Sodium acetate: prepare a solution of 3 M and pH 5.5 in deionized water. Adjust pH with acetic acid and not with HCl.
4. Solution #1 (for extraction of genomic DNA from a bacterial sample): 345 mM sucrose (Merck KGaA, Darmstadt, Germany), 10 mM Tris–HCl pH 8.0, 1 mM EDTA pH 8.0, 2 mg/mL lysozyme (Sigma-Aldrich Chemie GmbH, Munich, Germany).
5. Solution #2 (for extraction of genomic DNA from a bacterial sample): 300 mM NaCl, 2% (w/v) SDS, 100 mM Tris–HCl pH 8.0, 20 mM EDTA pH 8.0.
6. 2 mM DTT (1,4-dithiothreitol, stock: 1 M solubilized in deionized water, Carl Roth GmbH, Karlsruhe, Germany).
7. 50 μg/mL RNaseA.
8. Phenol–chloroform solution (ready to use): Roti-phenol–chloroform–isoamylalcohol (25:24:1), pH 7.5–8.
9. Chloroform–isoamylalcohol solution: Rotisolv chloroform mixed with Rotipuran isoamylalcohol (24:1).
10. TE buffer: 10 mM Tris–HCl, 1 mM EDTA, pH 8.0.
11. Loading buffer (6×): 50% (v/v) Rotipuran glycerol, 0.1% (w/v) SDS, 100 mM EDTA pH 8.0, 0.05% (w/v) Bromphenol blue.
12. DNA concentration is measured with a BioPhotometer (Eppendorf AG, Hamburg, Germany) in combination with the quartz cuvette TrayCell (Hellma Optik GmbH Jena, Jena, Germany) and a 1 mm or 0.2 mm lid.

2.5. Commercial Kits

1. Plasmid DNA preparation: innuPREP Plasmid Mini Kit (Analytik Jena Innuscreen GmbH, Jena, Germany) or NucleoBond Xtra Midi Kit (Macherey-Nagel GmbH & Co. KG, Dueren, Germany).
2. Agarose gel electrophoresis is performed according to a standard protocol and gels are stained with ethidium bromide (see also ref. 30, Vol. 1, Chap. 5).
3. DNA isolation from agarose gels: innuPREP Gel Extraction Kit (Analytik Jena Innuscreen GmbH, Jena, Germany).
4. Transposon mutagenesis: HyperMu MuA Transposase (Biozym Scientific GmbH, Hess. Oldendorf, Germany).

2.6. Enzymes

Enzymes are obtained from Fermentas GmbH (St. Leon-Rot, Germany) and applied with buffers at optimal reaction temperature.

1. Restriction enzymes: *Bam*HI (10 U/μL), *Bsp*143I (*Sau*3*A*I, 10 U/μL), *Bgl*II (10 U/μL), *Smi*I (*Swa*I, 10 U/μL).
2. SAP: shrimp alkaline phosphatase, 1 U/μL.
3. T4 DNA ligase, 1 U/μL.
4. T4 polynucleotide kinase, 10 U/μL.
5. Lysozyme from chicken egg white (Fluka, Sigma-Aldrich Chemie GmbH, Munich, Germany) dissolved in 10 mM Tris–HCl, 100 mg/mL, sterilized by sterile filtration (pore diameter of filter: 0.22 μm) and stored at −20°C.
6. Proteinase K from *Tritirachium album* (Merck KGaA, Darmstadt, Germany): solubilize in deionized water (10 mg/mL), sterilize by sterile filtration (pore diameter of filter: 0.22 μm), and store at −20°C.
7. DNase-free RNaseA is prepared according to ref. 30, Vol. 3, Appendix A4.39. Dissolve Ribonuclease A from bovine pancreas (Fluka, Sigma-Aldrich Chemie GmbH, Munich, Germany) at a concentration of 100 mg/mL in sodium acetate solution (0.01 M sodium acetate; pH of 5.2 is adjusted with to acetic acid). Heat to 100°C for 15 min. Cool slowly to room temperature. Adjust the pH by adding 0.1 volume of 1 M Tris–HCl (pH 7.4). Dispense into aliquots and store at −20°C.

2.7. DNA Ladder

1. GeneRuler 1 kb DNA ladder (Fermentas GmbH, St. Leon-Rot, Germany), 14 fragments (in bp): 10,000, 8,000, 6,000, 5,000, 4,000, 3,500, 3,000, 2,500, 2,000, 1,500, 1,000, 750, 500, 250.
2. 1 kb DNA extension ladder (Invitrogen GmbH, Karlsruhe, Germany), fragments (in bp): 40,000, 20,000, 15,000, 10,000, 8,144, 7,126, 6,108, 5,090/5,000, 4,072, 3,054, 2,026, 1,636, 1,010, 517/506.

2.8. Sequencing Primer of MuExpress

1. SeqA: 5′-ATCAGCGGCCGCGATCC-3′.
2. SeqB: 5′-TTATTCGGTCGAAAAGGATCC-3′.

3. Methods

3.1. Linearization and Dephosphorylation of the Shuttle Vector pEBP18

1. Isolate 50 μg of vector pEBP18 DNA from an overnight culture of *E. coli* DH5α (pEBP18) by using NucleoBond Xtra Midi Kit. Subsequently, determine DNA concentration photometrically.
2. Hydrolyze 50 μg of vector with 50 U *Bam*HI preferentially in a volume of 200–300 μL for 4 h at 37°C (see Note 1).

If the vector concentration is low, the reaction volume can be enhanced, but larger reaction volumes complicate subsequent DNA in gel purification. Thus, DNA concentration by isopropanol precipitation (see step 6) followed by elution in a smaller volume is recommended.

3. Analyze an aliquot (1–2 μL) of the hydrolyzed DNA and same amount of undigested vector as control by agarose gel electrophoresis to see if the vector is completely linearized and go directly to step 6.
4. If the vector is only partially digested, add 2.5 μL enzyme buffer (10×), 20 U *Bam*HI (2 μL), and 20.5 μL deionized water to the reaction vial, but ensure that the amount of *Bam*HI in the entire reaction does not exceed 5% (v/v). Vortex well. Incubate further for 2 h at 37°C.
5. Analyze an aliquot (2–3 μL) of the digested DNA by agarose gel electrophoresis to see if the vector is now completely linearized. Depending on the degree of digestion, continue with step 6 or repeat step 4.
6. Inactivate *Bam*HI by heat treatment (20 min at 80°C). Purify the vector by DNA precipitation with isopropanol by adding 0.1 volume 3 M sodium acetate and 0.7 volume isopropanol (−20°C) and mix carefully. Incubate the reaction mixture for 10 min on ice and subsequently centrifuge samples (16,000 × *g*, 30 min, RT). Discard the supernatant and wash the DNA pellet with 300 μL 70% (v/v) ethanol (−20°C) (see Note 2). Discard the supernatant and air-dry the DNA pellet.
7. Calculate the concentration of DNA ends based on 50 μg DNA (http://www.promega.com/biomath/):

$$\text{pmol DNA ends} = \mu\text{gDNA} \times \frac{\text{pmol}}{600\ \text{pg}} \times \frac{10^6\ \text{pg}}{\mu\text{g}} \times \frac{1}{N} \times 2 \times \frac{\text{kb}}{1{,}000\ \text{bp}}$$

"*N*" is the number of nucleotides (bp), "660 pg/pmol" is the average molecular weight of a single nucleotide pair, "2" is the number of ends in a linear DNA molecule, and "kb/1,000 bp" is a conversion factor for kilobases to base pairs.

8. Calculate dephosphorylation reaction with SAP according to Table 2 (see Note 3).
9. Resuspend the pellet in the estimated elution volume with 65°C deionized water and incubate the solution for 10 min at 65°C. Vortex carefully from time to time.
10. Cool sample down on ice, add SAP buffer, and SAP according to Table 2 and incubate for 60 min at 37°C.
11. Heat-inactivate SAP reaction by incubation at 65°C for 15 min. Continue with step 12 or store reaction at −20°C to continue plasmid preparation the next day.

Table 2
Composition of dephosphorylation reaction

Total reaction volume (μL)	50	50	50
SAP buffer (10×) (μL)	5	5	5
pmol DNA ends	1–10	10–20	20–25
SAP (1 U/μL) (μL)	1	2	2.5
Elution volume (μL)	44	43	42.5

12. Purify the dephosphorylated vector by agarose gel electrophoresis using a gel extraction kit (see Note 4).
13. Analyze the quality of vector preparation (i.e., efficiency of linearization and dephosphorylation) by a religation control assay. Set up three control reactions each containing 100 ng of the treated vector DNA in a total volume of 20 μL 1× T4 DNA ligation buffer. To the first reaction (reaction A), add 1 μL deionized water, to the second reaction (reaction B), add 0.5 μL T4 DNA ligase, and to the third reaction (reaction C), add 0.5 μL T4 DNA ligase and 0.5 μL T4 DNA polynucleotide kinase. After overnight incubation at 16°C transform chemical competent *E. coli* DH5α cells with 10 μL of the respective ligation mixture and estimate transformation efficiency.
14. Transformation efficiency of the three reactions is separately estimated by plating serial dilutions of transformants to obtain single colonies on selective agar plates followed by counting the colony forming units (cfu) after overnight incubation at 37°C. Transformation efficiency (cfu/μg DNA) can then be calculated:

$$\frac{\text{cfu}}{\mu\text{g}} = \text{cfu} \times \text{dilution factor} \times \frac{V_{\text{total}}}{V_{\text{plated}}} \times \frac{1{,}000\text{ng} / \mu\text{g}}{\text{ngDNA}}$$

"cfu": colony forming units estimated by counting bacterial colonies on selective agar plates, "V_{total}" is the volume of the whole transformation mixture (in μL or mL), "V_{plated}" is the volume of transformation mixture plated on selective agar plates (in μL or mL, same unit as "V_{total}"), "1,000 ng/μg" is a conversion factor for ng into μg, and "ng DNA" is the amount of DNA used for transformation.
15. Conclusion: reaction A (without T4 DNA ligase and polynucleotide kinase) illustrates the amount of circular nondigested vector, whereas reaction B (with T4 DNA ligase) gives information about the amount of digested but incompletely dephosphorylated vector molecules. Theoretically, transformation of (linear) plasmid DNA derived from reactions A and B

into *E. coli* should not lead to the formation of viable clones on the selective agar plates. Reaction C gives the total number of clones that can be obtained under the applied conditions. In comparison to control reactions A and B, the number of clones derived from reaction C should be significantly higher (>100×), otherwise the vector has to be further digested and/or dephosphorylated.

3.2. Isolation of Metagenomic DNA

In this section, a method will be described to isolate metagenomic DNA from a bacterial cell pellet derived from biofilms or lake water. This method should not be used to extract DNA from soil samples or sediments.

1. Estimate the weight of the cell pellet.
2. Gently resuspend ~1 g of cells in 3 mL solution #1 and incubate the sample in a water bath for 1.5 h at 37°C. During the following preparation, do not vortex the solution to avoid shearing of the DNA.
3. Mix the solution gently and periodically by inversion of the tube.
4. Add 6 mL solution #2 containing 2 mM DTT and 50 μg/mL RNaseA to each sample and incubate at 55°C for 30 min repeating step 3.
5. Add 100 μg/mL proteinase K and incubate at 55°C for 15 min repeating step 3.
6. Homogenize the DNA solution. Attach a cannula with a diameter of ~0.9 mm to the outlet of a syringe and aspirate DNA solution by pulling up the plunger. Afterwards, press the DNA solution slowly through the cannula with constant force and collect homogenate in a new falcon tube (see Note 5).
7. Remove proteins and purify DNA by extraction with phenol–chloroform (toxic, wear protective equipment, e.g., appropriate gloves, safety glasses, and use an extractor hood) according to ref. 30, Vol. 3, Appendix A8.9. The method is slightly modified and can briefly be summarized as follows.
   (a) Add 0.5–1 volume of phenol–chloroform solution to the homogenate.
   (b) Invert the tube until the solution gets turbid. Centrifuge the mixture (5 min, 3,000×*g*, 4°C) to separate the organic phase and aqueous phase again. If the phases are not properly separated, elongate centrifugation.
   (c) The organic phase is yellow colored. Usually, the aqueous DNA solution forms the upper phase. However, if the aqueous phase exhibits a high density because of increased salt (>0.5 M) or sucrose (>10%), concentrations, it may form the lower phase.

(d) Transfer the aqueous DNA phase into a new falcon tube without removing proteins from the interphase. Sometimes, it is helpful to use a modified pipette tip where the end is cut off to widen its opening.

(e) Attention: If the phenol–chloroform solution has not been adequately equilibrated to a pH of 7.8–8.0, the nucleic acids tend to migrate into the organic phase. Collect organic waste for disposal.

(f) Add 0.5–1 volume of chloroform–isoamylalcohol solution (toxic, wear protective equipment, e.g., appropriate gloves, safety glasses, and use an extractor hood) to aqueous phase and repeat step b.

(g) Transfer aqueous phase into several fresh 2 mL reaction tubes and discard organic waste.

8. Concentrate the DNA by isopropanol precipitation as describes before (see Subheading 3.1 step 6). After the addition of sodium acetate and isopropanol, invert the tube until DNA becomes visible as transparent to light white filaments. Store the solution at −20°C to adjourn isopropanol precipitation or continue with centrifugation. After washing the DNA with 70% (v/v) ethanol never over-dry precipitate, otherwise DNA cannot be solubilized again!
9. Dissolve the DNA pellet in 100 μL 65°C TE buffer by gently pipetting the solution up and down and incubate it for 10 min at 65°C. It is possible to extend the volume of TE buffer. Incubate the solution for several days but at least overnight at 4°C to allow the DNA to completely solubilize. Solutions containing dissolved high molecular weight DNA are usually viscous.
10. Determine the DNA concentration photometrically. If the DNA is not pure as seen by $A_{260/230}$ and $A_{260/280}$ values or contains traces of phenol (indicated by its characteristic smell), it has to be further purified. Usually, a second isopropanol precipitation step is sufficient.
11. Store the DNA for long-term storage at −20°C, otherwise at 4°C. Avoid numerous cycles of freezing and thawing.

3.3. Fragmentation of Metagenomic DNA

The fragmentation of metagenomic DNA is achieved by enzymatic digestion. For activity-based screenings, we recommend to generate "small" DNA fragments (up to 10 kb) from environmental DNA exhibiting a molecular weight of >40 kb. Proper fragmentation of metagenomic DNA is achieved upon partial hydrolysis by a restriction endonuclease which binds to 4-bp recognition sites. To facilitate cloning of DNA fragments, their ends should be compatible to a corresponding vector sites. Therefore, metagenomic DNA is digested using *Bsp*143I (*Sau*3AI isoschizomer) that generates *Bam*HI-compatible ends.

1. Analyze the quality of environmental DNA by agarose gel electrophoresis using a low percentage [0.5–0.6% (w/v)] TAE agarose gel (separation distance: 25–30 cm) and an adequate size marker (e.g., 1 kb DNA extension ladder). Agarose gel electrophoresis can be performed overnight applying low voltage (maximum 2 V/cm).
2. Conclusion: if the DNA is of high molecular weight (>40 kb) continue with step 3, otherwise repeat DNA isolation from the habitat sample. High content of RNA is indicated by a smear of low molecular weight fragments. RNA can be removed by specific hydrolysis using RNaseA followed by phenol–chloroform extraction of DNA.
3. For partial hydrolysis of metagenomic DNA, first prepare a premix containing 4 μg of isolated DNA and 4 μL enzyme buffer (10×) in a reaction volume of 36 μL. Incubate for 10 min at 37°C.
4. During incubation, pipette 2 μL loading buffer into seven 1.5 mL reaction tubes labeled 0, 2, 4, 6, 8, 10, 12 [min] and keep them on ice.
5. Add 4 μL *Bsp*143I (1 U/μL) to the tempered premix. Mix by pipetting up and down and transfer 4 μL of the reaction directly to a reaction tube (0 min) with loading buffer on ice. Incubate the residual reaction mixture at 37°C.
6. Transfer 4 μL aliquots from the reaction mixture every 2 min to the reaction tubes with loading buffer on ice, thereby generating seven aliquots of metagenomic DNA that were previously hydrolyzed with *Bsp*143I for 0, 2, 4, 6, 8, 10, and 12 min.
7. Analyze the restriction digest kinetics by agarose gel electrophoresis. Apply again a low percentage [0.5–0.6% (w/v)] TAE agarose gel, this time in combination with a separation distance of 10 cm and normal field strength.
8. Conclusion: the restriction digest should contain a mixture of variable size DNA fragments. This can be achieved only if digestion is neither complete (Fig. 4a) nor missing (Fig. 4b). Instead, a "DNA smear" of appropriate sized fragments should be obtained usually after 6–8 min of restriction enzyme digestion (Fig. 4c) with nondigested DNA still being visible.
9. Repeat restriction digest with reduced enzyme concentration until result is satisfying.
10. For preparative DNA fragmentation, run 10–15 reactions in parallel. Use a premix volume of 36 μL and incubate at 37°C for 10 min. Add 4 μL of *Bsp*143I solution (concentration estimated during kinetic study), incubate 6–8 min at 37°C, and stop the reaction by adding 8 μL loading buffer.

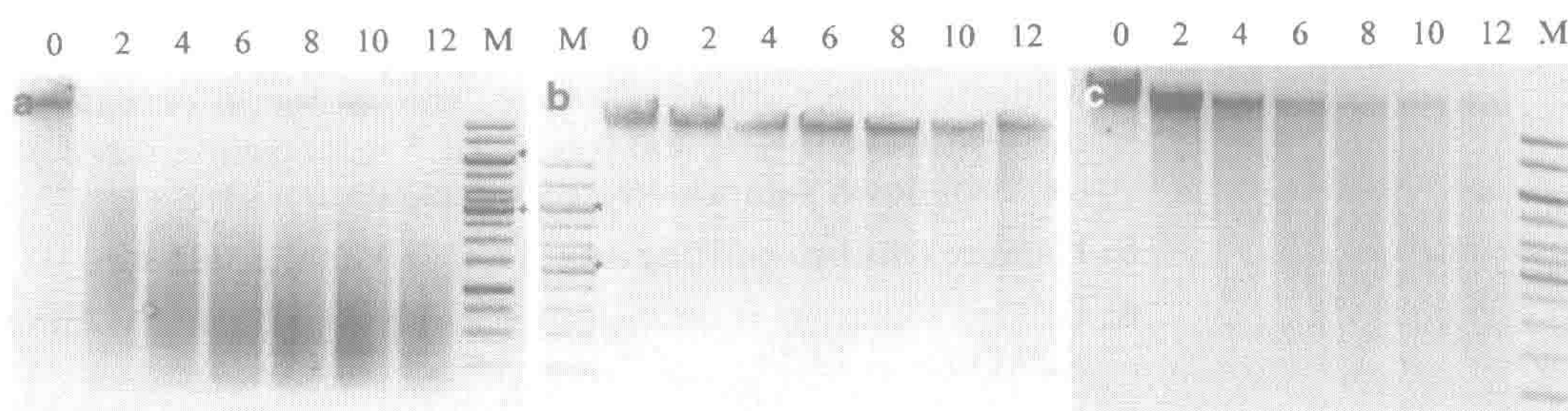


Fig. 4. Restriction endonuclease hydrolysis of metagenomic DNA digested with different amounts of restriction enzyme *Bsp*143I. Aliquots were taken after an incubation time of 0, 2, 4, 6, 8, 10, and 12 min, and fragment sizes were analyzed by agarose gel electrophoresis: (**a**) with 1 U *Bsp*143I/μg DNA nearly all DNA is degraded after 2 min; (**b**) with 0.1 U *Bsp*143I/μg DNA no degradation is observed even after 12 min; (**c**) with 0.5 U *Bsp*143I/μg DNA most of the environmental DNA appears as a smear with the expected size of 3–6 kb after 6–8 min. M: GeneRuler 1 kb DNA ladder (Fermentans), *asterisks* mark the 6 kb and *plus signs* the 3 kb fragment of the DNA standard.

11. To isolate hydrolyzed DNA fragments of appropriate size, a preparative gel electrophoresis is carried out: after electrophoresis, an agarose block containing the DNA fragments ranging from 3 to 6 kb is cut out from the gel and the DNA is subsequently purified by gel extraction using the innuPREP Gel Extraction Kit.
12. Analyze 200–300 ng of purified DNA fragments by gel electrophoresis to approve their correct size.

3.4. Ligation of Metagenomic DNA with Vector pEBP18

1. Calculate the concentration of vector and insert DNA (for insert DNA assume average insert size) (see Subheading 3.1, step 6.).
2. Calculate the amount of DNA needed for ligation based on 300–400 ng vector DNA and a ratio of vector to insert of 1:1 up to 5:1.
3. Set up ligation reaction in a total volume of 20 μL and 10 U T4 DNA ligase. Incubate overnight at 16°C.

3.5. Library Construction with the Shuttle Vector and Transformation of Bacterial Hosts

The preparation and transformation of electrocompetent *E. coli* cells were performed according to ref. 30, Vol. 1, Chap. 1, Protocol 26 with modifications.

3.5.1. Preparation of Electrocompetent E. coli Cells

1. Streak *E. coli* strain two times on LB agar plates and incubate overnight at 37°C to obtain viable and metabolic active cells.
2. Inoculate a single colony into 5 mL LB medium. Incubate the culture overnight at 37°C under constant shaking (120 rpm).
3. Inoculate 220 mL LB medium with an aliquot of the preculture (final OD_{600} = 0.05). Incubate cells at 37°C under permanent shaking (120 rpm). Determine the cell density of the growing culture (OD_{600}) until an optical density of OD_{600} = 0.4–0.5 is reached.

4. Collect 4 × 50 mL of cell suspension in 50 mL falcon tubes and incubate 20 min on ice. Use incubation time to cool down the centrifuge to 4°C and put 25 × 1.5 mL reaction tubes on ice; cool down solutions for cell preparation on ice.
5. Harvest cells by centrifugation (4°C, 10 min, 3,000 × *g*).
6. Discard supernatant und gently resuspend pellet in 1 volume MilliQ water.
7. Repeat steps 5 and 6, but this time resuspend the pellet in 0.5 volume MilliQ water. Pool the suspension from two falcons to obtain two new falcons with 50 mL suspension each.
8. Repeat steps 5 and 6 but resuspend cells in 1/50 volume glycerol solution (2 mL each). Pool the suspension from two falcons.
9. Repeat steps 5 and 6 but finally resuspend the pellet in 0.5 mL glycerol solution and aliquot samples of 25 µL. Incubate aliquots for 1 h at −20°C, then store at −80°C.

3.5.2. Transformation of Electrocompetent E. coli Cells with Test DNA

1. Thaw one aliquot of electrocompetent *E. coli* cells on ice and cool an electroporation cuvette (1 mm gap) on ice.
2. Add 1 µL pUC18 (20 ng/µL) to electrocompetent *E. coli* cells and mix gently by pipetting up and down.
3. Transfer suspension to electroporation cuvette; wipe electrodes on the outside of the cuvette with a paper towel to remove condensate. Place the cuvette in the electroporation device. Perform electroporation with Bio-Rad MicroPulser using program EC1 (1.8 kV, 5.8 ms).
4. Quickly resuspend cells in 600 µL EM1. Transfer solution to a reaction tube and incubate for 1 h at 37°C (120 rpm).
5. Estimate transformation efficiency (see Subheading 3.1, step 14).
6. Conclusion: transformation efficiency should be ~10^9/µg. If transformation efficiency is <10^8/µg, competent cells should be discarded.

3.5.3. Transformation of Electrocompetent E. coli Cells with Ligation Mixture

Metagenomic libraries constructed with the shuttle vector pEBP18 are routinely maintained in *E. coli* DH10B.

1. Transform electrocompetent *E. coli* DH10B with 1 µL ligation mixture as described in Subheading 3.5.2., but incubate for 3 h at 37°C (120 rpm) prior growth on medium supplemented with antibiotics. Plate serial dilutions of transformation mixture on selective EM1 agar plates and incubate overnight at 37°C. Store remaining transformation mixture at 4°C.
2. Calculate number of transformants on each plate to appraise optimal dilution factor.

3. Plate remaining transformation mixture according to step 2. If necessary, repeat transformation with ligation mixture to obtain >200,000 single colonies in total.
4. Cultivate 40 pEBP18 clones in 5 mL selective liquid LB medium and isolate plasmid DNA. In pEBP18, the insert is flanked by two *Smi*I recognition sites. Thus, hydrolyze recombinant pEBP18 DNA using *Smi*I (*Swa*I). Analyze restriction pattern by agarose gel electrophoresis.
5. Conclusion: hydrolysis of the empty pEBP18 vector generates two DNA fragments of 46 bp (hardly visible) and 10,660 bp. In contrast, the insert DNA of recombinant pEBP18 clones should be about 3–6 kb in size with inserts of 3–4 kb predominating over larger inserts. It is also possible that *Smi*I cuts within the inserts. In this case, estimate total insert size by adding the sizes of the respective insert fragments. The ratio of clones without an insert should not exceed 10%.
6. Transfer the clones with a sterile velvet stamp or a picking robot to agar plates with appropriate substrate to screen for enzyme activity. Alternatively, clones can be cultured directly on indicator plates. However, since colony forming units can vary when different media are applied, the number of transformants has to be calculated again.

3.5.4. Preparation of Electrocompetent P. putida KT2440 Cells and Transformation with Test DNA

An improved protocol for efficient transformation of *P. putida* KT2440 requires preparation of fresh competent *P. putida* KT2440 cells before transformation, since storage of competent *P. putida* KT2440 at −80°C reduces transformation efficiency by at least a factor of 10.

1. Streak *P. putida* KT2440 two times on EM1 agar plates. Incubate overnight at 30°C to obtain viable and metabolic active cells.
2. Inoculate a single colony into a vented Erlenmeyer flask with baffled bottom containing 20 mL EM medium. Incubate overnight at 30°C under constant shaking (120 rpm).
3. Transfer 6 × 2 mL of the overnight culture in 2 mL reaction tubes and harvest cells by centrifugation (16,000 × *g*, RT, 2 min). Wash pellets twice with 2 mL sucrose solution and resuspend cells in 600 μL sucrose solution. Pool the suspensions of three reaction tubes, harvest cells by centrifugation, and finally resuspend pellets in 100 μL sucrose solution to obtain a total volume of ~130 μL competent cells.
4. Mix competent cells with 1 μL of pBBR1MCS2 (20 μg/μL) and 1 μL ligation mixture (~15–20 ng vector), respectively. Transfer suspension to electroporation cuvettes (2 mm gap). Wipe electrodes on the outside of the cuvette with a paper

towel and place in the electroporation device. Perform electroporation with Bio-Rad MicroPulser using program EC2 (2.5 kV, 5.8 ms).

5. Quickly resuspend cells in 600 μL EM1 by pipetting up and down. Transfer solution into a test tube and incubate cells for 3 h at 30°C (120 rpm).
6. Plate serial dilutions of transformants on selective EM1 agar plates and estimate transformation efficiency of both reactions (see Subheading 3.1, step 14).
7. Conclusion: transformation efficiency with the control plasmid pBBR1MCS2 is typically ~10^7–10^8/μg. Transformation of ligation mixture usually reduces the efficiency by at least a factor of 10.
8. Cultivate 40 *P. putida* clones in 5 mL selective EM medium and isolate plasmid DNA. Analyze insert DNA by restriction digest with *Smi*I followed by agarose gel electrophoresis.
9. Conclusion: again, the insert DNA should be of 3–6 kb in size. The ratio of clones without insert should not exceed 10% (for more details see Subheading 3.5.3, step 5).

3.5.5. Library Construction in P. putida KT2440

In general, a metagenomic library should consist of ~200,000 clones carrying vector pEBP18 with diverse inserts. Therefore, calculate the amount of aliquots you need for electroporation.

1. Competent *P. putida* cells may be needed at a larger scale. For this, inoculate 5 mL EM medium with a single colony of *P. putida* KT2440 and grow the cells over day at 30°C under constant shaking (120 rpm). Inoculate an overnight culture with OD_{600} = 0.05 in a vented Erlenmeyer flask with baffled bottom and incubate cells at 30°C under shaking (120 rpm) for 16 h.
2. Collect cells by centrifugation. Depending on culture volume and reaction tube, the parameters vary: for 2 mL reaction tube use 16,000 × *g*, RT, 2 min; for 50 mL falcon tube use 3,000 × *g*, 4°C, 10 min (see Note 6).
3. Transform competent cells as described above (see Subheading 3.5, step 4).

3.5.6. Transformation of B. subtilis TEB1030

B. subtilis TBE1030 should be transformed utilizing its natural competence, because the shuttle vector has to integrate into the genome by homologous recombination at the amylase gene locus. *B. subtilis* TBE1030 can neither be transformed with ligation mixtures nor by protoplast transformation. Thus, single plasmids or plasmid pools obtained from the *E. coli* or *P. putida* libraries will be used.

1. Streak *B. subtilis* TEB1030 two times on EM1 agar plates. Incubate overnight at 37°C to obtain viable and metabolic active cells.
2. Inoculate a single colony within a vented Erlenmeyer flask with baffled bottom containing 5 mL Paris-Medium. Incubate cells overnight at 37°C and 120 rpm.
3. Inoculate 10 mL Paris-Medium in an Erlenmeyer flask with baffled bottom with 200 μL of the overnight culture. Determine cell density (OD_{600}) of the growing culture first after 2 h, then every 15 min. An optical density of $OD_{600} = 1$ is usually reached after 2.5–3 h.
4. Add 1 μg plasmid DNA to a 500 μL aliquot of the *B. subtilis* culture and further incubate the cells for 6 h at 37°C and 120 rpm. Use empty vector as a control.
5. Plate 200 and 300 μL samples of the transformed cells on selective EM1 agar plates and incubate overnight at 37°C. A few transformants should be obtained after 24–36 h.
6. Check amylase activity to test if transformation occurred via Campbell-type integration (single cross over) or by homologous recombination. Transfer colonies to a master plate and to a starch indicator plate and incubate overnight at 37°C. The next day, cover starch plate with 5 mL iodine solution, remove iodine solution after 2 min incubation. Conclusion: if homologous recombination has occurred, the amylase gene is disrupted and the colony appears black. If a halo appears around the colony, transformation has occurred via Campbell-type integration leaving amylase activity.
7. Correct transformants (Cm^R, Δ*amy*E) can be transferred on agar plates with substrate to analyze for enzyme activities.

3.6. Transposition with MuExpress

The transposon MuExpress can be applied to improve the expression efficiency for environmental genes in heterologous hosts as *E. coli* (26). MuExpress bases on the transposon MuA (33) that additionally harbors two divergently orientated T7 promoters. Thus, after random integration of MuExpress via nonspecific transposition, transcription of metagenomic genes may occur driven by T7 RNA polymerase.

3.6.1. Preparation of the Transposon MuExpress

Sufficient amount of plasmid pMuExpress can be obtained by standard plasmid preparation using NucleoBond Xtra Midi Kit from an overnight culture of *E. coli* DH5α (pMuExpress) grown on LB containing either kanamycin or chloramphenicol.

1. Cut 50 μg pMuExpress with 50 U *Bgl*II as described above (see Subheading 3.1, steps 1–4). The restriction digest with *Bgl*II delivers MuExpress fragments of 1,208 bp (Km^R) or

1,421 bp (Cm^R) depending on the variant of pMuExpress and the vector backbone of 2,984 bp.

2. Purify MuExpress by agarose gel electrophoresis with a gel extraction kit (check binding capacity of column, see Note 4).
3. Determine DNA concentration photometrically and analyze purity of preparation via agarose gel electrophoresis by loading 100 ng DNA.

3.6.2. In Vitro Transposition

The transposition of target DNA with the transposon MuExpress is carried out *in vitro*. The target DNA can vary and can consist of vector pools obtained from a plasmid or cosmid library as well as of single vector preparations.

1. Depending on the size of target DNA, the amount of DNA is different. Calculate the amount of target DNA per reaction. Use 300 ng DNA if the size of target is 10 kb or less, 600 ng DNA if the size of target is 20 kb, and up to 1,200 ng DNA if the size of target is 40 kb or larger.
2. Prepare the transposon insertion reaction mixture by adding components in the following order.
   (a) 2 μL HyperMu reaction buffer (10×).
   (b) *X* ng target DNA (see step 1).
   (c) 25 ng purified MuExpress transposon.
   (d) *Y* μL sterile deionized water to yield a reaction volume of 19 μL.
   (e) 1 μL HyperMu MuA Transposase (1 U/μL).
3. Incubate: 37°C, 2 h.
4. Stop the reaction by adding 1 μL HyperMu 10× stop solution. Mix and heat for 10 min at 70°C.

3.6.3. MuExpress Library Construction

1. Electrocompetent cells are prepared and transformed as described according to Subheading 3.5, steps 1–3 to construct a MuExpress library in *E. coli*.
2. The choice of host strain is dependent on the application. In general, the use of *recA*⁻ *endA*⁻ strains like DH10B or DH5α is preferable for target stability and subsequent purification steps, but not absolutely necessary. For transcription in both directions from T7 promoter, a bacterial strain harboring a T7 RNA polymerase gene like *E. coli* BL21(DE3) must be used.
3. Consider how many clones you will have to screen to cover DNA present in the original library by at least a factor of 2. It should be noted that a five- or higher fold coverage is preferable.
4. Resulting transformants are screened on selective agar plates containing the substrate of choice.

5. Isolate vector from "positive" transformants.
6. Retransform *E. coli* with the isolated vector and plate transformants again on selective agar plates containing the substrate of choice to confirm enzyme activity.
7. Identify sequence of active clones by sequencing the vector using transposon encoded primer sites SeqA and SeqB. (Further analysis of sequence and identification of target gene is not described.)

4. Notes

1. The restriction enzyme *Bam*HI exhibits star activity which appears at high enzyme concentrations and long incubation times. If *Bam*HI is used at low concentrations (1–2 U/μg DNA), usually no star activity occurs.
2. Do not vortex. It is more essential to remove salts from the wall of the reaction tube just by flipping the tube.
3. If higher concentrations are obtained than listed in Table 2, adjust the reaction volume according to this table.
4. Columns of commercial gel extraction kits usually only bind up to 10 μg of DNA. For gel extraction of 50 μg DNA, at least five columns must be used.
5. Homogenization is necessary to simplify the following phenol–chloroform extraction. Without homogenization, genomic DNA sticks to proteins, which accumulate in the interphase and removal of the aqueous phase might be difficult.
6. If cell preparation is carried out in a falcon tube, dilute the overnight culture with MilliQ water before performing the first centrifugation step. Stationary grown *P. putida* cells usually stick to each other and are difficult to resuspend. Therefore, use 25 mL overnight culture and add 25 mL MilliQ water.

References

1. Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M. (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5, R245–R249.
2. Curtis, T.P. and Sloan, W.T. (2005) Microbiology. Exploring microbial diversity – a vast below. *Science* 309, 1331–1333.
3. Ferrer, M., Golyshina, O., Beloqui, A., and Golyshin, P.N. (2007) Mining enzymes from extreme environments. *Curr Opin Microbiol* 10, 207–214.
4. van den Burg, B. (2003) Extremophiles as a source for novel enzymes. *Curr Opin Microbiol* 6, 213–218.
5. Brennan, Y., Callen, W.N., Christoffersen, L., Dupree, P., Goubet, F., Healey, S., et al. (2004) Unusual microbial xylanases from insect guts. *Appl Environ Microbiol* 70, 3609–3617.
6. Entcheva, P., Liebl, W., Johann, A., Hartsch, T., and Streit, W.R. (2001) Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from

microbial consortia. *Appl Environ Microbiol* 67, 89–99.

7. Knietsch, A., Bowien, S., Whited, G., Gottschalk, G., and Daniel, R. (2003) Identification and characterization of coenzyme B_{12}-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. *Appl Environ Microbiol* 69, 3048–3060.
8. Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A., and Daniel, R. (2003) Construction and screening of metagenomic libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. *Appl Environ Microbiol* 69, 1408–1416.
9. Daniel, R. (2005) The metagenomics of soil. *Nat Rev Microbiol* 3, 470–478.
10. Steele, H.L., Jaeger, K.-E., Daniel, R., and Streit, W.R. (2009) Advances in recovery of novel biocatalysts from metagenomes. *J Mol Microbiol Biotechnol* 16, 25–37.
11. Green, B.D. and Keller, M. (2006) Capturing the uncultivated majority. *Curr Opin Biotechnol* 17, 236–240.
12. Ferrer, M., Beloqui, A., Timmis, K.N., and Golyshin, P.N. (2009) Metagenomics for mining new genetic resources of microbial communities. *J Mol Microbiol Biotechnol* 16, 109–123.
13. Galvão, T.C., Mohn, W.W., and de Lorenzo, V. (2005) Exploring the microbial biodegradation and biotransformation gene pool. *Trends Biotechnol* 23, 497–506.
14. Schloss, P.D. and Handelsman, J. (2003) Biotechnological prospects from metagenomics. *Curr Opin Biotechnol* 14, 303–310.
15. Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 68, 669–685.
16. Lämmle, K., Zipper, H., Breuer, M., Hauer, B., Buta, C., Brunner, H., et al. (2007) Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. *J Biotechnol* 127, 575–592.
17. Cottrell, M.T., Moore, J.A., and Kirchman, D.L. (1999) Chitinases from uncultured marine microorganisms. *Appl Environ Microbiol* 65, 2553–2557.
18. Handelsman, J. (2005) Sorting out metagenomes. *Nat Biotechnol* 23, 38–39.
19. Gabor, E., Liebeton, K., Niehaus, F., Eck, J., and Lorenz, P. (2007) Updating the metagenomics toolbox. *Biotechnol J* 2, 201–206.
20. Streit, W.R. and Schmitz, R.A. (2004) Metagenomics – the key to the uncultured microbes. *Curr Opin Microbiol* 7, 492–498.
21. Gabor, E.M., Alkema, W.B., and Janssen, D.B. (2004) Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ Microbiol* 6, 879–886.
22. Ferrer, M., Martínez-Abarca, F., and Golyshin, P.N. (2005) Mining genomes and 'metagenomes' for novel catalysts. *Curr Opin Biotechnol* 16, 588–593.
23. Li, Y., Wexler, M., Richardson, D.J., Bond, P.L., and Johnston, A.W.B. (2005) Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of *Rhizobium leguminosarum* and of *Escherichia coli* reveals different classes of cloned *trp* genes. *Environ Microbiol* 7, 1927–1936.
24. Martinez, A., Kolvek, S.J., Yip, C.L., Hopke, J., Brown, K.A., MacNeil, I.A., et al. (2004) Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Appl Environ Microbiol* 70, 2452–2463.
25. Wang, C., Meek, D.J., Panchal, P., Boruvka, N., Archibald, F.S., Driscoll, B.T., et al. (2006) Isolation of poly-3-hydroxbutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants. *Appl Environ Microbiol* 72, 384–391.
26. Leggewie, C., Henning, H., Schmeisser, C., Streit, W.R., and Jaeger, K.-E. (2006) A novel transposon for functional expression of DNA libraries. *J Biotechnol* 123, 281–287.
27. Schmeisser, C., Stöckigt, C., Raasch, C., Wingender, J., Timmis, K.N., Wenderoth, D.F., et al. (2003) Metagenome survey of biofilms in drinking-water networks. *Appl Environ Microbiol* 69, 7298–7309.
28. Nelson, K.E., Weinel, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Martins dos Santos, V.A., et al. (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ Microbiol* 4, 799–808.
29. Eggert, T., Brockmeier, U., Dröge, M.J., Quax, W.J., and Jaeger, K.-E. (2003) Extracellular lipases from *Bacillus subtilis*: regulation of gene expression and enzyme activity by amino acid supply and external pH. *FEMS Microbiol Lett* 225, 319–324.
30. Sambrook, J. and Russell, D.W. (2001) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Press, New York.
31. Kovach, M.E., Elzer, P.H., Hill, D.S., Robertson, G.T., Farris, M.A., Roop, R.M., et al. (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* 166, 175–176.
32. Cronan, J.E. (2003) Cosmid-based system for transient expression and absolute off-to-on

transcriptional control of *Escherichia coli* genes. *J Bacteriol* 185, 6522–6529.

33. Haapa, S., Taira, S., Heikkinen, E., and Savilahti, H. (1999) An efficient and accurate integration of mini-Mu transposons in vitro: a general methodology for functional genetic analysis and molecular biology applications. *Nucleic Acids Res* 27, 2777–2784.

Chapter 9

Screening of Functional Promoter from Metagenomic DNA for Practical Use in Expression Systems

So-Youn Park and Geun-Joong Kim

Abstract

The functional overproduction of proteins is an essential step required for providing enzymes and proteins for practical applications. Accordingly, the expression system used is important, e.g., *cis*-acting elements including promoters for the expression of recombinant proteins in a broad range of hosts or preferential expression in a specific host. We have introduced a bidirectional promoter trap system (pBGR1) for screening of promoters with up- or/and down regulatory elements from resources of metagenomic DNA. The pBGR1 is equipped with a pair of fluorescent proteins acting as reporters, facing opposite directions without promoters; this is a promising system that can function regardless of the direction of inserted promoters. Subsequently, promoters trapped by pBGR1 can be employed in the construction of new expression vectors for practical applications.

Key words: Promoter, Dual reporter, Metagenome, Promoter trap, Expression system

1. Introduction

Potential *cis*-acting promoter elements – both inducible and noninducible – have recently attracted tremendous attention. Following the unveiling of the genome, various research projects have focused on elucidating genomic information; along with combinatorial molecular biological engineering techniques, these have resulted in a rapid expansion of the sequence universe of promoters (1, 2). However, several difficult problems remain to be solved before the functional overexpression of versatile proteins originating from various organisms, including uncultivable strains, can be achieved. This is primarily because most expression vectors that are currently available function only in a limited range of hosts and also require an expensive inducer. Many attempts are therefore being conducted to screen and select useful expression vectors

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668, DOI 10.1007/978-1-60761-823-2_9, © Springer Science+Business Media, LLC 2010

from varied biological resources (3–6). In this context, microbial resources that have so far remained difficult to culture comprise a metagenome that would be a treasure trove for finding useful promoters and/or regulatory proteins for protein expression.

Metagenomes are being continuously searched for finding useful genetic materials with a primary focus on the screening and identification of open reading frames encoding for enzymes and proteins (7). Although there are few examples that specifically indicate how the regulatory element might be exploited in practical applications, such instances have led us to expect that the extremely diverse traits caused by evolutionary environmental pressures can allow the metagenome to possess a wide variety of regulatory elements, including promoters, in all organisms.

In general, the trapping and identification of promoters is performed by inserting a genomic DNA fragment into the upstream region of a reporter protein (fluorescent, luminescent, or chromogenic protein), thereby selecting the trapped promoter by the expression of the reporter (4, 8, 9). However, this kind of promoter-trap system has certain drawbacks, i.e., several reporter proteins need to be supplemented by expensive inducers and have low sensitivity and high background noise. The fundamental problem is that only promoters trapped in the same orientation as the reporter gene are detected. Thus, a trapping method that is easier, more efficient, and more rapid in screening and identifying promoters from various resources is urgently required.

Currently, functional promoters in cells are readily screened from genetic resources using promoter trapping systems that utilize a bidirectional probe concept. For example, a pair of reporters, GFP and *DsRed*, has been employed by inserting them in the opposite direction without promoters, which then functioned regardless of the direction of inserted genes, including the *cis*-acting regulatory element in a host (5).

2. Materials

2.1. Strain and Medium

1. *Escherichia coli* XL1-Blue (endA1 gyrA96 (nalR) thi-1 recA1 relA1 lacglnV44 F'[::Tn10 proAB+ lacIqΔ(lacZ)M15] hsdR17 (rK-mK+)) from Stratagene (see Note 1).
2. Competent *E. coli* cells are cultured in SOB medium (20 mM $MgCl_2$) and prepared by using Inoue transformation buffer [$MnCl_2·4H_2O$ 55 mM, $CaCl_2·2H_2O$ 15 mM, KCl 250 mM, PIPES (0.5 M, pH 6.7) 10 mM] (see Note 2).
3. Medium: Luria–Bertani (LB) medium is commonly used for cell culture. Terrific Broth (TB) and M9 medium are used for screening of functional promoters when required.

2.2. Plasmids

1. pBluescript II-SK(+) (Stratagene) is used as a template for the construction of trap-vector systems of the pBGR series. pDsRed2-N1 (Clontech, USA) and pGFPuv (Clontech) are used as the sources of red and green fluorescent proteins, respectively.
2. Commercially available pBluescript II-SK(+), pTrc99a (Pharmacia), and pMAL-c2x (New England Biolabs) plasmids along with two plasmids, pQE-1767 (10) and pSmGlu (11), are used in the comparison of protein expression.

2.3. DNA Extraction and Purification

1. Plasmids are to be purified by using the Wizard Plus SV Minipreps, and DNA recovery from agarose gel slices is conducted by using Wizard SV Gel & PCR Clean-up System (Promega).
2. The soil used for extracting the metagenome is directly frozen in liquid nitrogen, and then kept in a deep freezer at −80°C. Total metagenomic DNA is extracted using the FastDNA® SPIN Kit for Soil (Bio 101, USA) (see Note 3).

2.4. Protein Expression and Cell Lysis

1. Isopropyl β-D-1-thiogalactopyranoside (IPTG) is dissolved in distilled H_2O and sterilized by passing through a 0.22-μm disposal filter. Store the solution at −20°C.
2. Lysis buffer (20 mM Tris–HCl pH 8.0) is mainly used for cell lysis. The protease inhibitor cocktail solution (Sigma) and Triton X-100 should be supplemented as and when needed.

2.5. Polyacrylamide Gel Electrophoresis

1. PAGE gel: Acryl/Bis™ 37.5:1 (30:0.8), 40% (w/v) solution (Amresco), 1.5 M Tris–HCl pH 8.8, 1 M Tris–HCl pH 6.8, 10% SDS, 10% APS, and *N,N,N,N′*-tetramethylethylenediamine (TEMED, Sigma).
2. Tris–glycine electrophoresis buffer (1×): 25 mM Tris, 250 mM glycine pH 8.3, 0.1% (w/v) sodium dodecyl sulfate (SDS). Store the buffer at room temperature.
3. SDS gel-loading buffer (1×): 50 mM Tris–HCl pH 6.8, 100 mM dithiothreitol (DTT), 2% (w/v) SDS, 0.1% bromophenol blue, 10% (v/v) glycerol. Store at −80°C. (DTT and SDS are excluded from the gel-loading buffer when preparing the native gel-loading buffer).
4. Activity staining for esterase: α-naphthyl acetate (45 mg/mL) dissolved in ethoxyethanol and Fast Blue RR (15 mg/mL) dissolved in dimethyl sulfoxide (DMSO) (see Note 4). Store at −20°C. Activity staining for glucosidase: 5 mM 4-methylumbelliferyl β-D-glucopyranoside (MUG) dissolved in DDW or DMSO; this is to be prepared fresh before use.

3. Methods

For promoter trapping, a bidirectional vector system with a pair of reporter proteins (green and red fluorescent proteins) and a cloning site (*Bam*HI) between them on a suitable backbone of a plasmid, designated pBGR1, which is highly sensitive and reproducible, is to be used. This system can select the trapped promoter by detecting the expression of the reporter protein by cloning the metagenomic DNA onto the cloning site *Bam*HI. The frequency of positive hits and the screening time, however, vary according to the size of the insert and the maturation time of the reporter proteins, especially of the red fluorescent protein. As reported in another study (4), the trapping system of pBGR1 is partly similar to the SIGEX procedure of operon screening in the case of inducible proteins, although it can screen the catabolic operon, including promoters, by reporting a reporter in a direction. Therefore, the screening procedure using pBGR1 mainly focuses upon the screening of constitutively expressible promoters. The procedure is detailed below.

3.1. Construction of a Bidirectional Fluorescence Reporter System for Promoter Trapping

1. Amplify the *GFPuv* and *DsRed* genes, using forward and reverse primer sets designed according to the respective coding regions of pDsRed2-N1 and pGFPuv, by using PCR to construct a trap plasmid. In the case of the *GFPuv* coding gene, the *Bam*HI site used for the cloning site in the trap vector is naturally present in the gene and needs to be removed. This is readily achieved by PCR-based site-directed mutagenesis and is designated GFP_{bk}.
2. Digest the vector pBluescript II-SK(+) and *DsRed* gene with the two restriction enzymes, *Hin*dIII and *Eco*RI, respectively, that are intentionally incorporated into the primer sequences. Separate the appropriate fragments by electrophoresis through an agarose gel and then purify the digested DNA by Wizard SV Gel and PCR Clean-up System. In this step, pBluescript II-SK without T3 and T7 promoters by digestion with *Pvu*II can also be used. Incubate the purified vector and *DsRed* gene with ligase overnight at 4°C or for 4 h at 16°C. Transform the competent *E. coli* with the ligation mixture and then purify the recombinant plasmid designated pBRI.
3. Digest the *pBRI* and GFP_{bk} gene with the restriction enzyme *Xba*I (incorporated into both the primers used for PCR) and construct the pBGR1 vector as described in step 2 above (see Note 5). The plasmid map of pBGR1 is shown in Fig. 1.
4. Identify the pBGR1 construct using sequence analysis and check the reporter protein expression without a trapped promoter using an ultraviolet (UV) hand lamp and/or fluorescent imaging analyzer (see Note 6).

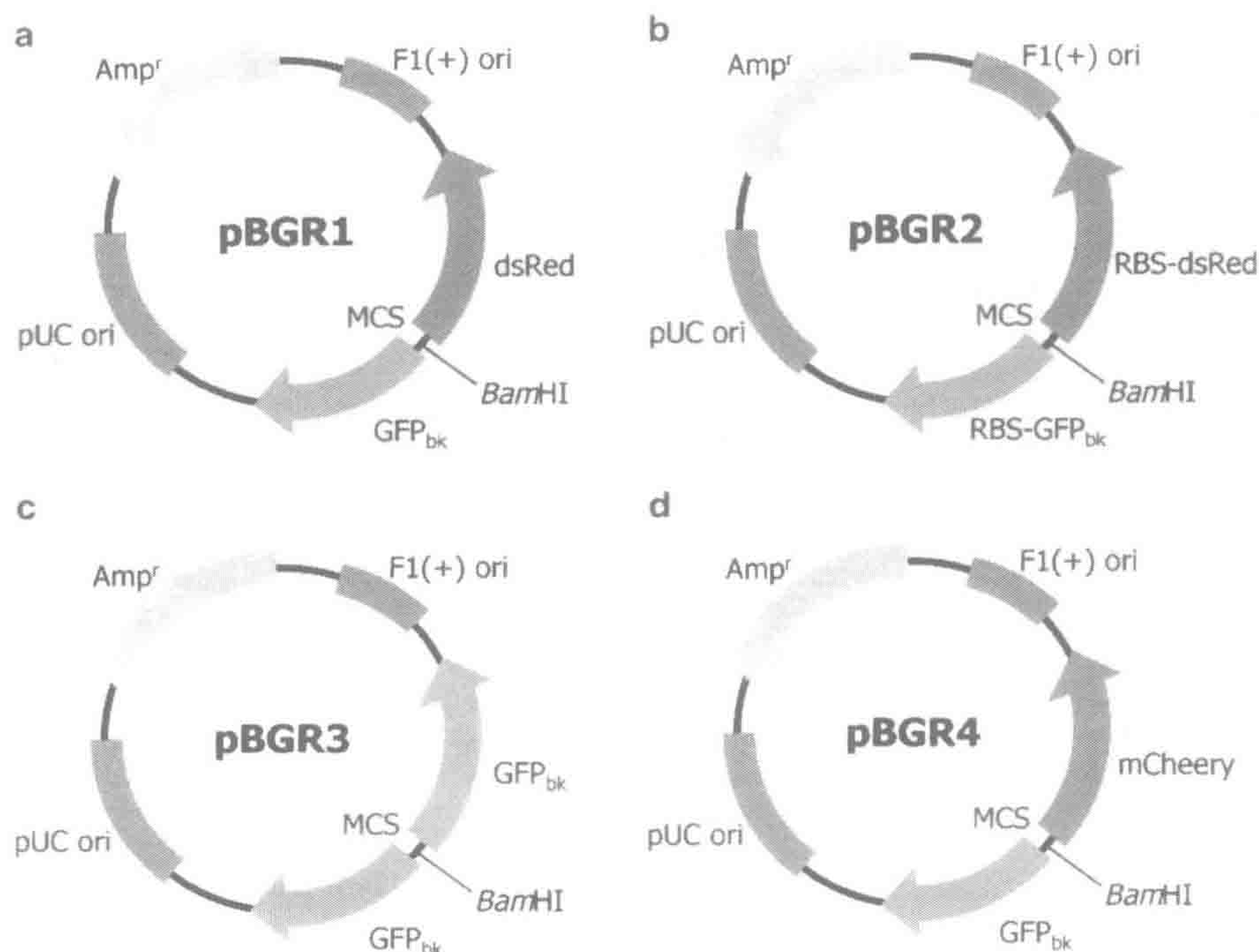


Fig. 1. Schematic representation of the genetic map of promoter-trap pBGR vector. Two reporters equipped in each trap vector are as follows: (**a**) GFP_{bk} and *dsRed*; (**b**) GFP_{bk} and *dsRed* with its own *RBS*; (**c**) GFP_{bk} and GFP_{bk}; (**d**) GFP_{bk} and *mCheery*. A series of pBGR vectors can be constructed with suitable components to the user's objectives.

3.2. Preparation of Metagenomic DNA

1. Slowly thaw the soil at 4°C.
2. Extract the genomic DNA from soil by using the FastDNA® SPIN Kit for soil. At this time, avoid severe shearing force, which decreases the positive hits due to random fragmentation of genetic resources.
3. Identify the diversity of the purified metagenomic DNA by restriction fragment length polymorphism with certain restriction enzymes, or by 16s rRNA sequence analysis of the gene amplified by PCR using a set of universal primers (12, 13).

3.3. Construction and Screening of Metagenomic Library for Functional Promoter

1. Prepare the competent cells by the commonly used procedure (14). Inoculate 1 mL of SOB medium containing 20 mM $MgCl_2$ with a single colony of *E. coli* XL1-Blue. After cultivation at 37°C for 12 h, the cells are reseeded into 30–100 mL of SOB medium containing 20 mM $MgCl_2$. Grow the cells for 2.5 h at 37°C and then wash the harvested cells by using the transformation buffer containing 20 mM hexamine cobalt chloride. Store the competent cells at –80°C.
2. Purify the *Bam*HI-digested pBGR1 vector. Digest the metagenomic DNA with one of the restriction enzymes *Sau*3AI or *Bam*HI. Separate the digested DNA by electrophoresis

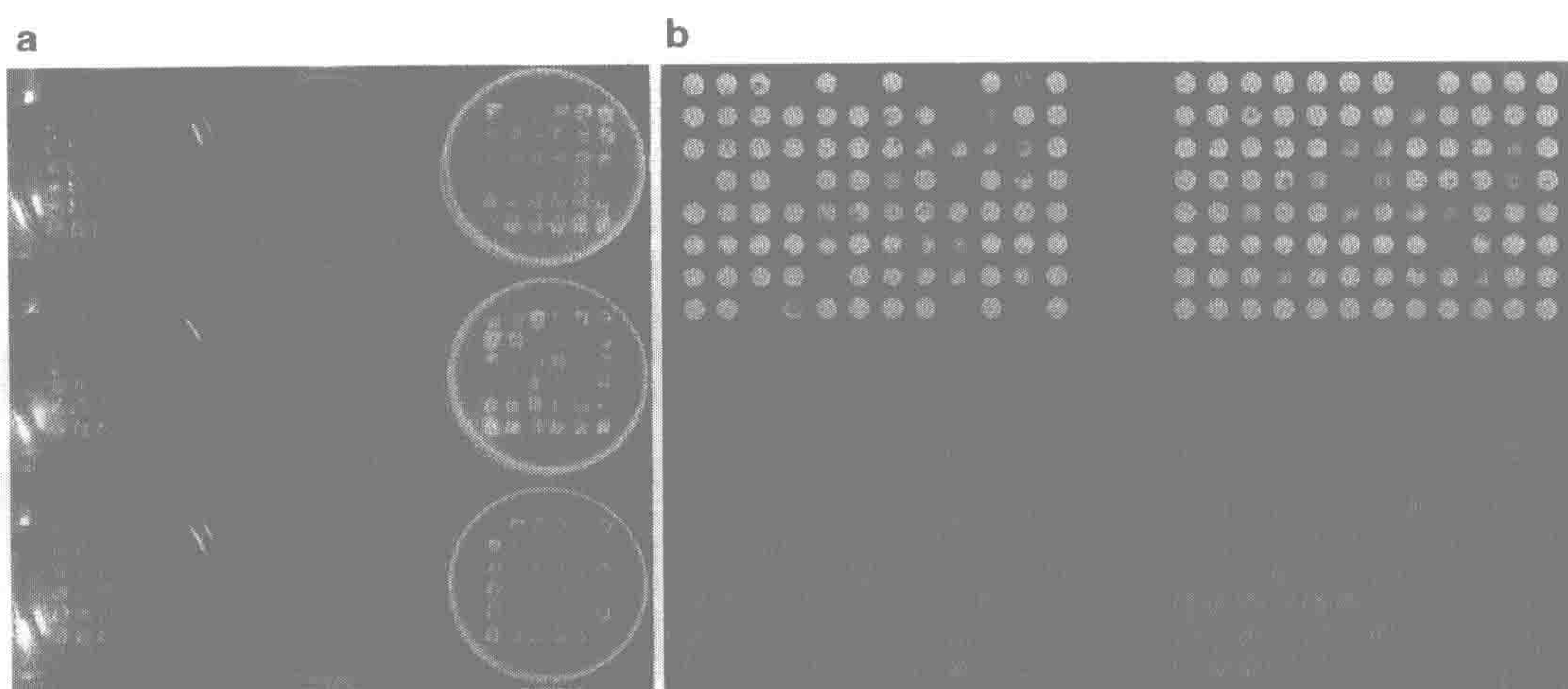


Fig. 2. Construction of metagenomic DNA library with pBGR1 system. A typical recombinant *E. coli* library is primary screened on solid plates (**a**) and further screened based on the fluorescence emitted from a format of HTS culture (**b**). In this case, a 96-well plate is used.

through an agarose gel and then purify the DNA fragments of 200–2,000 bp (see Note 7).

3. Mix the digested pBGR1 vector and the insert DNA, and incubate the mixture with ligase overnight at 4°C. Transform the competent *E. coli* cells with the ligation mixture and culture the transformant onto the LB agar medium (50 μg/mL ampicillin) at 37°C.
4. Check and screen the fluorescent promoter-trapped clones in the library by using a UV hand lamp or fluorescence imaging analyzer [GFP is excited at 395 nm, and *DsRed* is excited at 558 nm (15, 16)] after 20–24 h incubation at 37°C. At this time, use the positive control (pDsRed2-N1 and pGFPuv) and the negative control (empty vector), respectively. An example of a typical recombinant library is shown in Fig. 2.
5. Screen the stable promoters from the primary screened clones using specific conditions, varying the culture temperature as well as the component of medium (see Note 8).

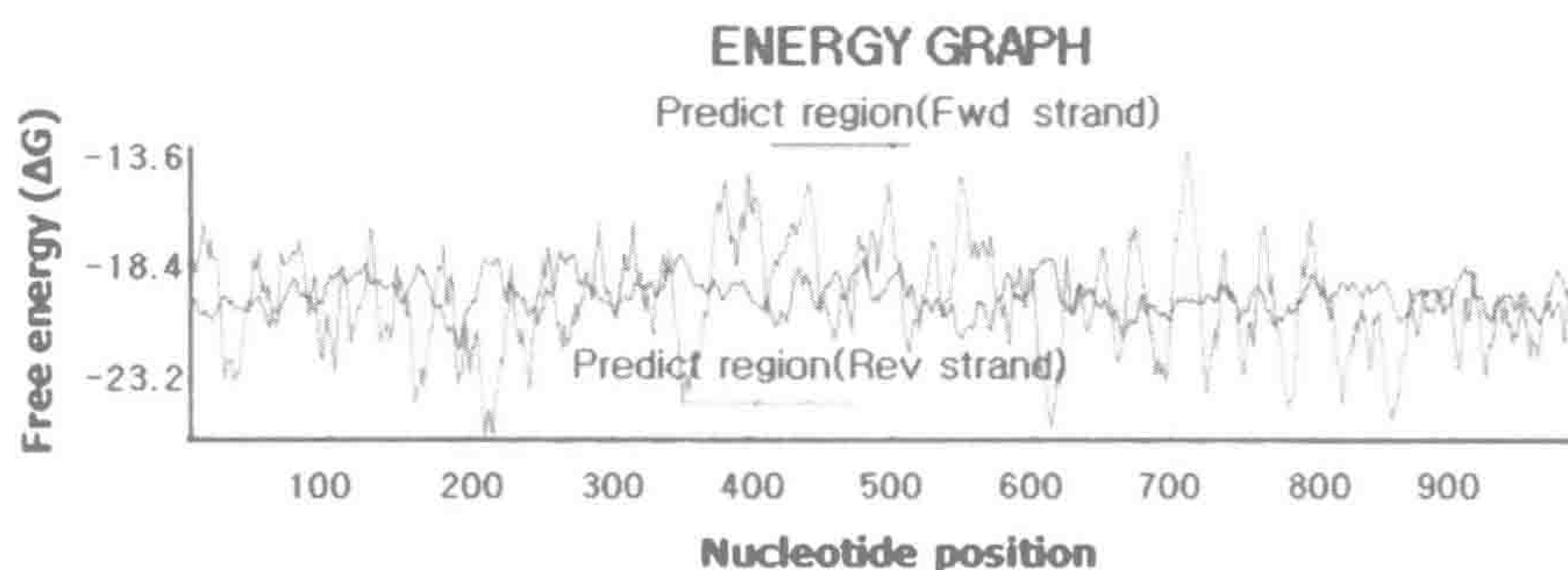
3.4. Analysis of Expression of Reporter Proteins in the Clone with Trapped Genes

1. Cultivate the final screened promoter-trap clones in 5 mL of LB medium (50 μg/mL ampicillin) for 10 h at 37°C and harvest by centrifugation (5,000 ×*g*, 5 min), and then store the pellets at –80°C.
2. Repeat freezing and thawing twice, and resuspend the cells in 500 μL of 20 mM Tris–HCl buffer pH 8.0. Lyse the cells by sonication, immediately centrifuge for 30 min at 4°C and then transfer the supernatant to a fresh tube.
3. Add 3 μL of 5× SDS or native gel-loading buffer to 15 μL aliquots of all samples and load 15 μL of the samples on a

10% SDS-PAGE and 8% Native-PAGE. Analyze the reporter protein expression by examining fluorescence bands on native PAGE using a UV hand lamp (see Note 9). The expression levels of reporter proteins are calculated from the width and density of the corresponding band on SDS-PAGE.

3.5. Analyzes of Structural Organization and Putative Promoters in the Trapped Genes

1. Amplify the trapped gene (spanned by both reporter genes) by using PCR with a set of primers (arbitrarily select a sequence from both reporter genes) and analyze the nucleotide sequence.
2. Compare the nucleotide sequence of the trapped gene as a query using the default option of BLASTX/BLASTN (NCBI, http://blast.ncbi.nlm.nih.gov/Blast.cgi). Predict the promoter region by using software provided by Neural Network Promoter Prediction (http://promotor.biosino.org/) and Prom-find (http://nucleix.mbu.iisc.ernet.in/prompredict/) (see Note 10 and Fig. 3).
3. Align the sequences of the putative promoter region and then compare these with the defined consensus of various



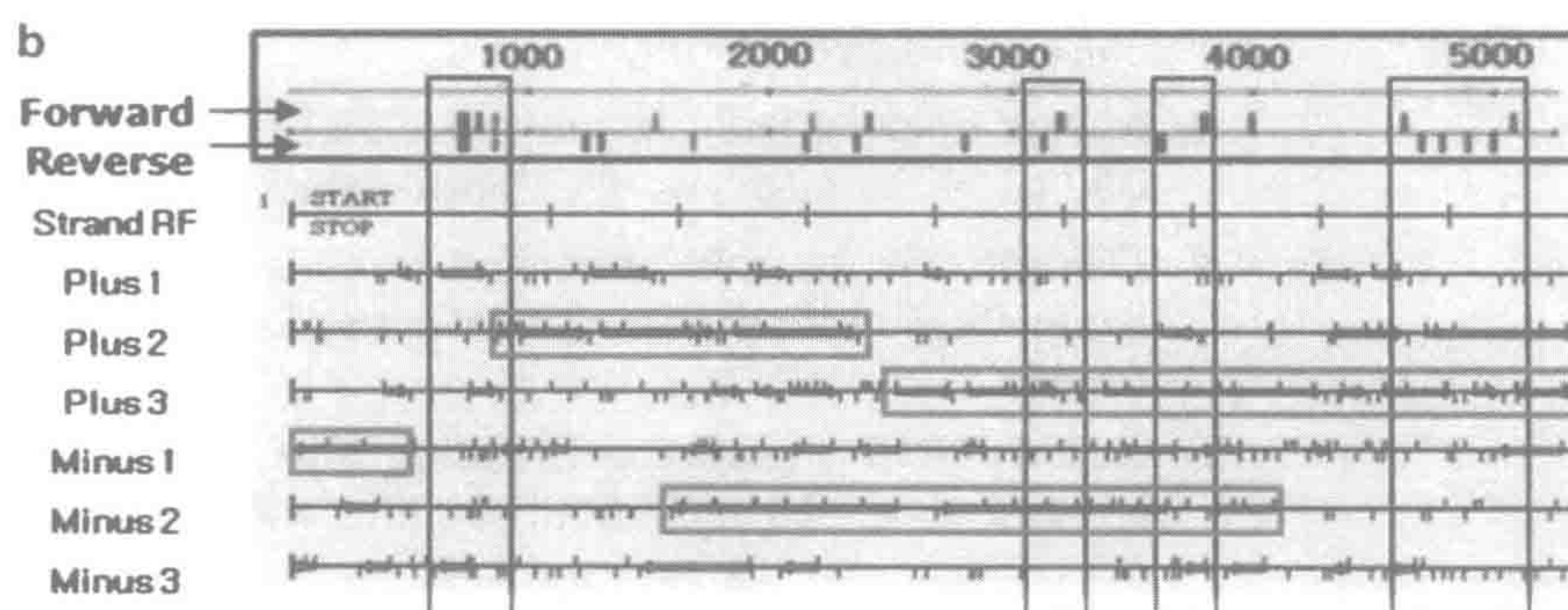


Fig. 3. Prediction results of a putative promoter in selected gene fragments. (a) Promoter of the unknown gene is predicted by using Prom-find, which is based on DNA stability. (b) Promoter of the nucleotide sequence is predicted by a combination of the program NNPP and the ORF finder. The regions that overlap in the results of each program provide the probable regions for the promoter.

promoters from a cell (in this case, use the consensus of the promoter from *E. coli*) by using a multiple sequence alignment program (e.g., ClustalW).

4. Verify the promoter function by a DNase footprinting assay (17) or RT PCR (18).

3.6. Construction of Expression Vector Equipped with Trapped Promoters

1. Select the predicted or verified promoters containing either typical or weak signal sequence patterns of the transcription element (−35, −10, +1) (see Note 10). Amplify the promoter region by PCR using a primer containing a ribosome-binding site and multicloning site (arbitrarily designed based on the physical map of the DNA fragment of the amplified promoter gene).
2. Insert the amplified promoter into the promoterless pBluescript II-SK(+) and transform the resulting construct into competent cells.
3. Isolate the new expression vectors equipped with a novel promoter from the metagenome, and then confirm the organization by DNA sequencing.
4. Insert a readily detectable reporter, such as *GFPuv*, β-glucosidase, or esterase gene, into the downstream region of the promoter and then evaluate the expression ability of the new construct. The same genes are also cloned into commercially available expression vectors with their own respective promoters, i.e., pTrc-99a (trc promoter), pMAL-c2x (tac promoter), and pBluescript II-SK(+) (T3, T7 promoter) vector (see Note 11).
5. Inoculate a single colony of each construct described above into 5 mL of LB (50 μg/mL ampicillin) medium and culture for 24 h at 37°C. Add 1 mM IPTG into the culture of control group constructs with an inducible promoter and culture further for 2 h. Collect identical amounts of cells from each culture and store at −80°C.
6. Repeat freezing and thawing twice and resuspend the cells in 250 μL of 20 mM Tris–HCl buffer pH 8.0.
7. Disrupt cells by sonication and immediately centrifuge the lysate for 30 min at 4°C. Transfer the supernatant to a fresh tube.
8. Add 5× SDS or native gel-loading buffer to an appropriate amount (e.g., 10 μg) of all samples and load it onto 10% SDS-PAGE and 8% Native-PAGE. Analyze the total protein expression patterns and solubility by SDS-PAGE.
9. Analyze the enzyme function and appropriate folding by resolving the band on native PAGE, i.e., fluorescence of *GFPuv* with UV excitation. For evaluating the staining activity of esterase, soak the native PAGE gel with 40 mL of

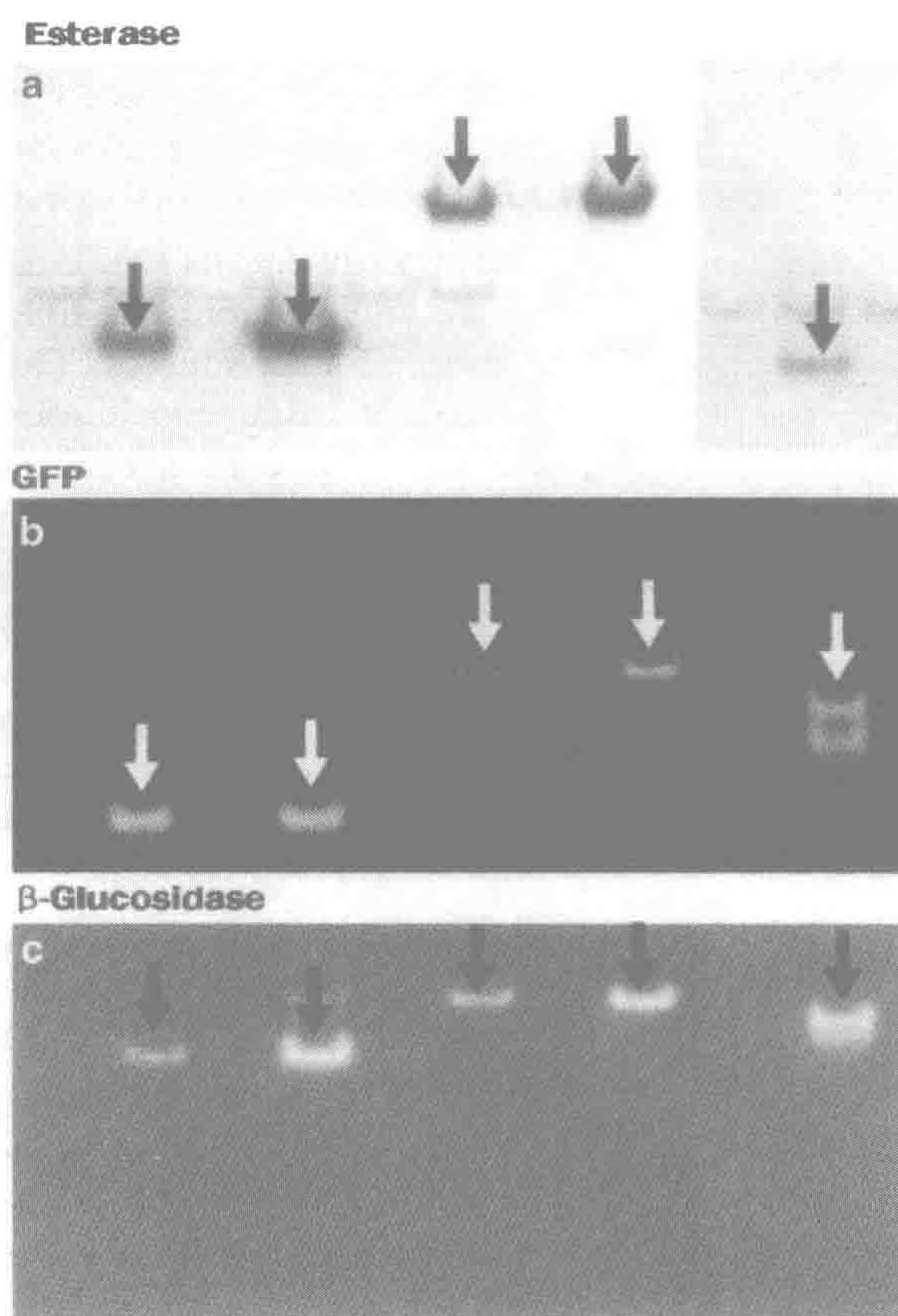


Fig. 4. Staining activity for evaluation of a new expression vector. A new construct is evaluated by protein expression that is readily detected on native PAGE in vitro (esterase, GFP, and β-glucosidase).

50 mM Tris–HCl pH 7.0 buffer, add 150 μL of α-naphthyl acetate (45 mg/mL), and then add 300 μL of Fast Blue RR (15 mg/mL). Monitor the rapidly developing (<10 min) brown-colored band. For evaluating the staining activity of glucosidase, the 0.8–1% agar pH 5.0–7.0 containing 5 mM of MUG is applied to the native PAGE gel; analyze the fluorescence band with UV excitation (345 mM) after a short incubation time of approximately 5–10 min (see Note 12). An example of zymogram is shown in Fig. 4.

4. Notes

1. *E.coli* XL1-Blue is a commonly used strain for the library construction and screening of genetic resources. If you want to screen a constitutive or inducible promoter in a specific host, you can choose any host that can be genetically manipulated.

2. The frequency of positive hits for promoter screening is closely linked with the competence of the host cell and also the library size that results from a single shot of transformation with the appropriate amount of ligation mixture (50–100 μg).
3. The strategy of total DNA extraction depends upon the source of the environmental sample; it is difficult to apply a single principle for all sources. The metagenomic DNA extraction process is being continuously optimized for perfecting an individual or combined process (including enzymatic lysis, bead-beating, and chemophysical method) (19). However, there is no generalized protocol for the extraction of genomes from soils containing complex physiochemical elements. The purity and diversity of genomes are a basis for the efficiency of library construction and must be considered in the extraction process. Therefore, various attempts to increase the yield and diversity are recommended (20).
4. Fast Blue RR is used for esterase detection, interacting with the ester hydrolysis product and then changes its color to dark brown. When the Fast Blue RR is spontaneously oxidized to deep brown from brown in DMSO, it cannot be used for staining.
5. A series of pBGR vectors can be constructed with suitable components according to the user's objectives. The pBGR1 vector described in this work can screen the promoter mainly by translational fusion of the reporter protein. Insertion of RBS into the upstream region of a reporter or of both genes can occasionally efficiently screen functional promoters due to transcriptional fusion with the reporter gene. A pBGR vector construct with mCherry (a monomeric mutant of *DsRed*) (21) as the pair for GFPbk can also be applied for screening functional promoters in a short time. When needed, a pair of reporters can be incorporated into a broad range of plasmids, such as pBBR122.
6. You must verify that there is no expression of the reporter protein without a trapped promoter. Hence, a transformed cell with an empty vector should be subjected to excitation, and the fluorescence should be checked according to the time course.
7. Metagenomic DNA fragments are recovered from the size-fractionated agarose gel. If you want to screen the operon, including functional promoter or *cis*-acting elements, larger-sized DNA fragments – measuring approximately 3–10 kb – are used. At this time, pBGR2 with its own RBS is used as a trap vector for promoter screening.
8. A constitutively expressible promoter under a specific condition can be screened by modulation of culture conditions (with an inducer or physical shock). In addition, an inducible promoter can also be screened according to methods described by other studies (22, 23).

9. When the expression of fluorescent protein is analyzed by SDS-PAGE, occasionally, there is no protein band of the innate size. This is because various sizes of reporter proteins can result from fusion with the inserted gene.
10. Two approaches are currently available for promoter prediction. The first involves evaluating the consensus of sequence alignment, based on comparison with a known promoter sequence of various organisms using the software NNPP or BLAST (Fig. 3). The other approach is the thermodynamic approach that predicts probable regions of DNA that melt readily (24). To obtain a high fidelity from promoter prediction, we strongly suggest that the two approaches be combined appropriately. Occasionally, a promoter with a weak signal with no typical consensus sequence results in a high expression of proteins.
11. Staining activity on the solid plate makes it easy to select the clones expressing esterase or β-glucosidase. Positive clones with esterase are selected rapidly (<10 min) by overlaying a soft agar (0.6%) supplemented with α-naphthyl acetate (45 mg/mL) and Fast Blue RR (15 mg/mL). However, false positives result from long incubation periods (>10 min) due to the background activity of various esterases of host cells. Positive clones with β-glucosidase are selected based on the fluorescence by overlaying a 0.6% soft agar supplemented with 2.5 mM 4-MUG. Adequate cooling to prevent spontaneous reaction is considered necessary. Cell culture is conducted by adding inducer (IPTG) for protein expression as and when needed.
12. Excess protein loading on the PAGE gel may result in a smeared band during staining.

Acknowledgments

This work was partly supported by the research grant (M10866020003-08N6602-00311) from the Korean Ministry of Education, Science and Technology through the Korean Science and Engineering Foundation.

References

1. Bajic, V.B., Tan, S.L., Suzuki, Y., and Sugano, S. (2004) Promoter prediction analysis on the whole human genome. *Nat Biotechnol* 22, 1467–1473.
2. Hartner, F.S., Ruth, C., Langenegger, D., Johnson, S.N., Hyka, P., Lin-Cereghino, G.P., et al. (2008) Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. *Nucleic Acids Res* 36, e76.
3. Chen, S., Bagdasarian, M., Kaufman, A.G., and Walker, E.D. (2007) Characterization of strong promoters from an environmental

Flavobacterium hibemum strain by using a green fluorescent protein-based reporter system. *Appl Environ Microbiol* 73, 1089–1100.

4. Uchiyama, T., Abe, T., Ikemura, T., and Watanabe, K. (2005) Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat Biotechnol* 23, 88–93.
5. Han, S.S., Lee, J.Y., Kim, W.H., Shin, H.J., and Kim, G.J. (2008) Screening of promoters from metagenomic DNA and their use for the construction of expression vectors. *J Microbiol Biotechnol* 18, 1634–1640.
6. Dunn, A.K. and Handelsman, J. (1999) A vector for promoter trapping in *Bacillus cereus*. *Gene* 226, 297–305.
7. Cowan, D., Meyer, Q., Stafford, W., Muyanga, S., Cameron, R., and Wittwer, P. (2005) Metagenomic gene discovery: past, present and future. *Trends Biotechnol* 23, 321–329.
8. Rasko, D.A., Esteban, C.D., and Sperandio, V. (2007) Development of novel plasmid vectors and a promoter trap system in *Francisella tularensis* compatible with the pFLN10 based plasmids. *Plasmid* 58, 159–166.
9. Frey, P.M., Scharer-Hernandez, N.G., Futterer, J., Potrykus, I., and Puonti-Kaerlas, J. (2001) Simultaneous analysis of the bidirectional African cassava mosaic virus promoter activity using two different luciferase genes. *Virus Genes* 22, 231–242.
10. Kim, J.H., Choi, G.S., Kim, S.B., Kim, W.H., Lee, J.Y., Ryu, Y.W., et al. (2004) Enhanced thermostability and tolerance of high substrate concentration of an esterase by directed evolution. *J Mol Catal B Enzym* 27, 169–175.
11. Kim, J.Y., Lee, J.Y., Shin, Y.S., and Kim, G.J. (2009) Mining and identification of a glucosidase family enzyme with high activity toward the plant extract indican. *J Mol Catal B Enzym* 57, 284–291.
12. Fennell, D.E., Rhee, S.K., Ahn, Y.B., Haggblom, M.M., and Kerkhof, L.J. (2004) Detection and characterization of a dehalogenating microorganism by terminal restriction fragment length polymorphism fingerprinting of 16S rRNA in a sulfidogenic, 2-bromophenol-utilizing enrichment. *Appl Environ Microbiol* 70, 1169–1175.
13. Dickie, I.A. and FitzJohn, R.G. (2007) Using terminal restriction fragment length polymorphism (T-RFLP) to identify mycorrhizal fungi: a methods review. *Mycorrhiza* 17, 259–270.
14. Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* 166, 557–580.
15. Waldo, G.S., Standish, B.M., Berendzen, J., and Terwilliger, T.C. (1999) Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol* 17, 691–695.
16. Bevis, B.J. and Glick, B.S. (2002) Rapidly maturing variants of the Discosoma red fluorescent protein (DsRed) (vol 20, pg 83, 2002). *Nat Biotechnol* 20, 1159–1159.
17. Leonetti, J.P., Wong, K., and Geiduschek, E.P. (1998) Core-sigma interaction: probing the interaction of the bacteriophage T4 gene 55 promoter recognition protein with *E.coli* RNA polymerase core. *EMBO J* 17, 1467–1475.
18. Lee, H.J., Jeon, H.J., Ji, S.C., Yun, S.H., and Lim, H.M. (2008) Establishment of an mRNA gradient depends on the promoter: an investigation of polarity in gene expression. *J Mol Biol* 378, 318–327.
19. Frostegard, A., Courtois, S., Ramisse, V., Clerc, S., Bernillon, D., Le Gall, F., et al. (1999) Quantification of bias related to the extraction of DNA directly from soils. *Appl Environ Microbiol* 65, 5409–5420.
20. Roh, C., Villatte, F., Kim, B.G., and Schmid, R.D. (2006) Comparative study of methods for extraction and purification of environmental DNA from soil and sludge samples. *Appl Biochem Biotechnol* 134, 97–112.
21. Shaner, N.C., Steinbach, P.A., and Tsien, R.Y. (2005) A guide to choosing fluorescent proteins. *Nat Methods* 2, 905–909.
22. Gao, B., Flores, S.C., Bose, S.K., and McCord, J.M. (1996) A novel *Escherichia coli* vector for oxygen-inducible high level expression of foreign genes. *Gene* 176, 269–270.
23. Cardona, S.T. and Valvano, M.A. (2005) An expression vector containing a rhamnose-inducible promoter provides tightly regulated gene expression in *Burkholderia cenocepacia*. *Plasmid* 54, 219–228.
24. Kanhere, A. and Bansal, M. (2005) A novel method for prokaryotic promoter prediction based on DNA stability. *BMC Bioinformatics* 6, 1.

Chapter 10

Substrate-Induced Gene Expression Screening: A Method for High-Throughput Screening of Metagenome Libraries

Taku Uchiyama and Kentaro Miyazaki

Abstract

The SIGEX (substrate-induced gene expression) method is a novel approach for the screening of gene (genome) libraries. In addition to the commonly used function- and sequence-driven approaches to screening, SIGEX provides a third option; in SIGEX, positives are identified using a reporter gene, and the library is constructed using an "operon-trap" vector. This vector contains the reporter gene immediately downstream of the cloning site for the genomic insert so that the expression of the inserted gene(s) is coupled with that of the reporter gene. This system is especially suitable for screening catabolic genes that are induced in response to metabolically relevant compounds, such as substrates. If expression of the inserted gene(s) is activated in response to the addition of these compounds, then positive clones can be identified based on the reporter signal. The most effective selection is obtained by the use of a FACS (fluorescence-activated cell sorter) in conjunction with a FACS-compatible fluorescent reporter protein, such as GFP (green fluorescent protein). Activity-based screening of metagenomic libraries often suffers from low sensitivity and low throughput. In contrast, the high throughput, high sensitivity, and versatility of SIGEX make it a particularly suitable method for screening metagenomic libraries.

Key words: FACS, Flow cytometer, GFP, Operon-trap vector, Metagenome library, SIGEX, Screening

1. Introduction

As vast reservoirs of genetic and metabolic diversity, the bacterial species have great potential value for various biotechnological applications. However, recent research in molecular microbial ecology has shown that the great majority of bacterial species are difficult to pure-cultivate in the lab by means of conventional microbiological techniques (1, 2). Counteracting the technological limitations of cultivation-based techniques, direct cloning of

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668, DOI 10.1007/978-1-60761-823-2_10,

environmental DNA (known as the "metagenome") is anticipated to lead to breakthroughs in the discovery process (3–5).

Two approaches for exploring the rich genetic resource provided by bacteria have been commonly used: activity-based (functional) screening and nucleotide sequence-based screening. Although these approaches have been used successfully and are well established, they are generally labor-intensive and time-consuming, and they suffer from problems of low throughput and low sensitivity. To overcome these drawbacks, we developed a third screening method, designated SIGEX (substrate-induced gene expression) (6–8). A schematic outline of the SIGEX screening procedure is shown in Fig. 1.

In the SIGEX scheme, a metagenomic library is constructed using an "operon-trap" vector. This vector contains a reporter gene immediately downstream of the insertion site for cloning and thus couples the expression of inserted genes to the expression of the reporter gene. If expression of the insert is activated, positive clones can be identified by the signal derived from the

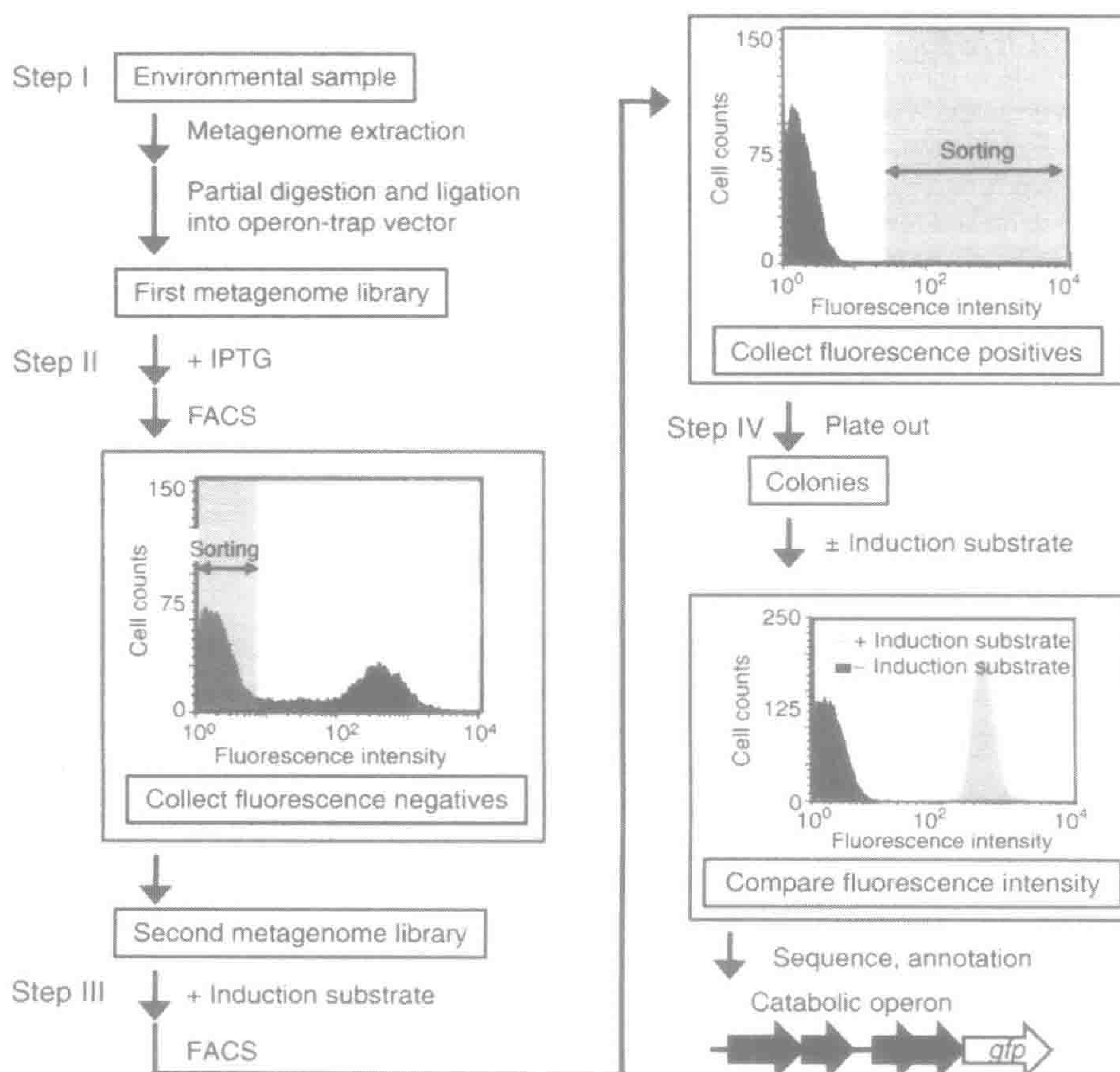


Fig. 1. Schematic representation of the SIGEX screening procedure.

reporter gene product. The selection process is most efficient when performed using a FACS (fluorescence-activated cell sorter) and a FACS-compatible fluorescent protein, such as green fluorescent protein (GFP), as a reporter. Using this approach, ultrahigh-throughput screening can be achieved (30,000 clones/s).

SIGEX was first developed to screen for catabolic genes that are induced in response to metabolically related compounds, such as the specific substrates and metabolites of the gene(s) of interest. Since the expression of catabolic genes is often regulated by metabolically related compounds, catabolic genes and their substrate-responsive genetic elements, such as transcriptional regulators, are good targets for identification by SIGEX. However, because SIGEX does not actually assay enzymatic activities, the interpretation of SIGEX results is not always straightforward. The analysis of "positive" clones may lead to surprising identifications, but these unexpected results sometimes lead to the discovery of novel biological events or genes. A typical experimental protocol for SIGEX is described below.

2. Materials

2.1. Cell Culture

1. LB medium: Dissolve 10 g Bacto-peptone (BD Diagnostics), 5 g Bacto-yeast extract (BD Diagnostics), and 5 g NaCl in 950 mL distilled water. Adjust pH to 7.3 by adding HCl, and bring to 1 L with distilled water. Sterilize by autoclaving, and store at 4°C.
2. SOC medium: Dissolve 20 g Bacto-peptone, 5 g Bacto-yeast extract, 0.58 g NaCl, and 0.19 g KCl in 950 mL distilled water. Adjust pH to 7.0 by adding HCl, bring to 970 mL with distilled water, and sterilize by autoclaving. Cool to room temperature, and add 10 mL of sterile 1 M $MgCl_2$ (sterilized by autoclaving) and 20 mL of sterile 1 M glucose [sterilized by passage through a 0.22-μm filter (Millipore)]. Store at 4°C.
3. dLB medium: Dissolve 1 g Bacto-tryptone (BD Diagnostics), 0.5 g Bacto-yeast extract, 1 g NaCl, and 2 g maltose in 950 mL distilled water. Adjust pH to 7.3 by adding HCl, bring to 990 mL with distilled water, and sterilize by autoclaving. Cool to room temperature and add 10 mL of 1 M $MgSO_4$ (sterilized by autoclaving). Store at 4°C.
4. Ampicillin stock solution: Dissolve ampicillin in distilled water at 100 mg/mL. Pass the resulting solution through a 0.22-μm filter (Millipore) to sterilize, and store at 20°C.

5. Selective media: Add ampicillin stock solution to LB or dLB medium to a suitable final concentration (i.e., 100 μg/mL). Store at 4°C for up to 1 month.
6. LB agar plates: Add 15 g Bacto agar (BD Diagnostics) to 1 L of LB medium. Sterilize by autoclaving, cool to 60°C, and add ampicillin stock solution to a final concentration of 100 μg/mL. Immediately dispense solution into Petri dishes (94 × 16 mm; Greiner). Store at 4°C for up to 1 month.
7. Isopropyl β-D-1-thiogalactopyranoside (IPTG) stock solution: Dissolve IPTG in distilled water at a final concentration of 0.1 M. Pass the resulting solution through a 0.22-μm filter (Millipore) to sterilize, and store at −20°C.
8. Incubator/rotary shaker that can be used at various temperatures and speeds.

2.2. Metagenomic DNA Preparation

1. 1 M Tris–HCl pH 8.0: Dissolve 121.1 g of tris (hydroxymethyl) -aminomethane (Tris) in 800 mL of distilled water. Adjust pH to 8.0 by adding HCl, and bring to 1 L with distilled water. Sterilize by autoclaving, and store at room temperature.
2. 0.5 M EDTA: Add 186.1 g of $Na_2EDTA \cdot 2H_2O$ to 800 mL distilled water and adjust pH to 8.0 by adding NaOH pellets. (The disodium salt of EDTA will not dissolve until pH of the solution is adjusted to ~pH 8.0.) Adjust the volume to 1 L with distilled water, sterilize by autoclaving, and store at room temperature.
3. 10% Hexadecyltrimethylammonium bromide (CTAB): Dissolve 100 g of CTAB in 900 mL of distilled water and heat to 60°C to aid dissolution. Adjust the volume to 1 L with distilled water, and store at room temperature.
4. 20% SDS: Dissolve 200 g of sodium dodecyl sulfate (SDS) in 900 mL distilled water. Stir at 60°C to aid dissolution. When the SDS is completely dissolved, adjust the volume to 1 L with distilled water. Store at room temperature.
5. DNA extraction buffer: Dissolve 13.5 g of Na_2HPO_4, 0.64 g of NaH_2PO_4, and 87.7 g of NaCl in 500 mL of distilled water, adjust the volume to 600 mL with distilled water, and sterilize by autoclaving. Cool to room temperature and add 100 mL of 1 M Tris–HCl pH 8.0, 200 mL of 0.5 M EDTA, and 100 mL of 10% CTAB. Store at room temperature for up to 1 week.
6. Lysozyme solution: Dissolve 0.2 g of chicken egg white lysozyme (Wako) in 10 mL of 0.25 M Tris–HCl pH 8.0 immediately before use.

7. Proteinase K solution: Dissolve 0.2 g of *Tritirachium album* proteinase K (Wako) in 10 mL of 0.25 M Tris–HCl pH 8.0 immediately before use.
8. 3 M sodium acetate pH 5.2: Dissolve 408.3 g of $CH_3COONa·3H_2O$ in 800 mL of distilled water. Adjust pH to 5.2 with glacial acetic acid. Adjust the volume to 1 L with distilled water. Sterilize by autoclaving, and store at room temperature.
9. TE (Tris-EDTA)-saturated phenol (Nippon Gene) (see Note 1).
10. Phenol/chloroform/isoamyl alcohol (PCI; Nippon Gene) (see Note 2).
11. TE buffer: Mix 1 mL of 1 M Tris–HCl pH 8.0 and 0.2 mL of 0.5 M EDTA pH 8.0 in 90 mL of distilled water. Adjust the volume to 1 L with distilled water. Sterilize by autoclaving, and store at room temperature.
12. Incubator/rotary shaker that can be used at various temperatures and speeds.
13. Centrifuge.

2.3. Metagenomic Library Preparation

1. DNA ligation kit LONG (TaKaRa).
2. Alkaline phosphatase solution, from shrimp (Wako).
3. *Bam*HI (TaKaRa).
4. *Sau*3AI (TaKaRa).
5. RECOCHIP (TaKaRa).
6. Competent cells *E. coli* JM109 (TaKaRa).
7. DNA electrophoresis equipment.
8. UV transilluminator (365 nm).
9. Electroporator (MicroPulser; Bio-Rad).
10. Centrifuge.

2.4. Flow Cytometer

1. Flow cytometer: FACS Vantage SE (BD Biosciences) or equivalent. The flow cytometer should be equipped with a laser for excitation at 488 nm. A 530/30-band-pass filter is used for measuring GFP fluorescence. A FACS sorting unit that can be used to separate GFP-expressing cells from non-fluorescent cells is needed. All data analysis can be performed using Cellquest Pro software (BD Biosciences).
2. Calibration beads for flow cytometer: Fluoresbrite calibration grade 6.0-µm YG microspheres (Polysciences).
3. Phosphate-buffered saline (PBS): Dissolve 8 g NaCl, 0.2 g KCl, 1.1 g Na_2HPO_4, and 0.2 g KH_2PO_4 in distilled water. Adjust pH to 7.4 by adding HCl, and adjust the volume to 1 L with distilled water. Pass through a 0.22-µm filter and sterilize by autoclaving. Store at room temperature.

3. Methods

The SIGEX scheme (Fig. 1) consists of four steps.

Step 1 is the construction of a metagenomic library. The metagenome is partially digested and ligated into an operon-trap vector (e.g., p18GFP; Fig. 2) that is then used to transform host cells, most commonly *E. coli*. The operon-trap vector includes a coexpressed reporter gene such as *gfp*, which encodes GFP. The GFP fluorescence is used to detect clones that contain metagenomic fragments with genes that are expressed in response to induction substrates. As a cloning host, we usually use *E. coli*, but other organisms may also be used. We have found some host-derived bias in screening; SIGEX tends to select for metagenomic fragments from original hosts (although the hosts' identities are unknown) that are phylogenetically related to the cloning host. Using a bacterial species other than *E. coli* as a host might be advantageous in avoiding this inherent host bias.

Step 2 is to remove "false-positive" clones that constitutively express GFP in the absence of inducer substrate. The library is grown in the absence of substrates relevant to the genes of interest and then subjected to FACS screening. In our system, we add IPTG to efficiently remove the background; clones that fluoresce in response to IPTG probably contain self-ligated vectors (no insert), which can be a major source of background fluorescence during screening. Some clones containing metagenomic inserts might include constitutively expressed genes, and some researchers might be interested in trapping such genetic elements. However, if the intent is to screen for substrate-dependent activation, these constitutively expressed genes should be eliminated.

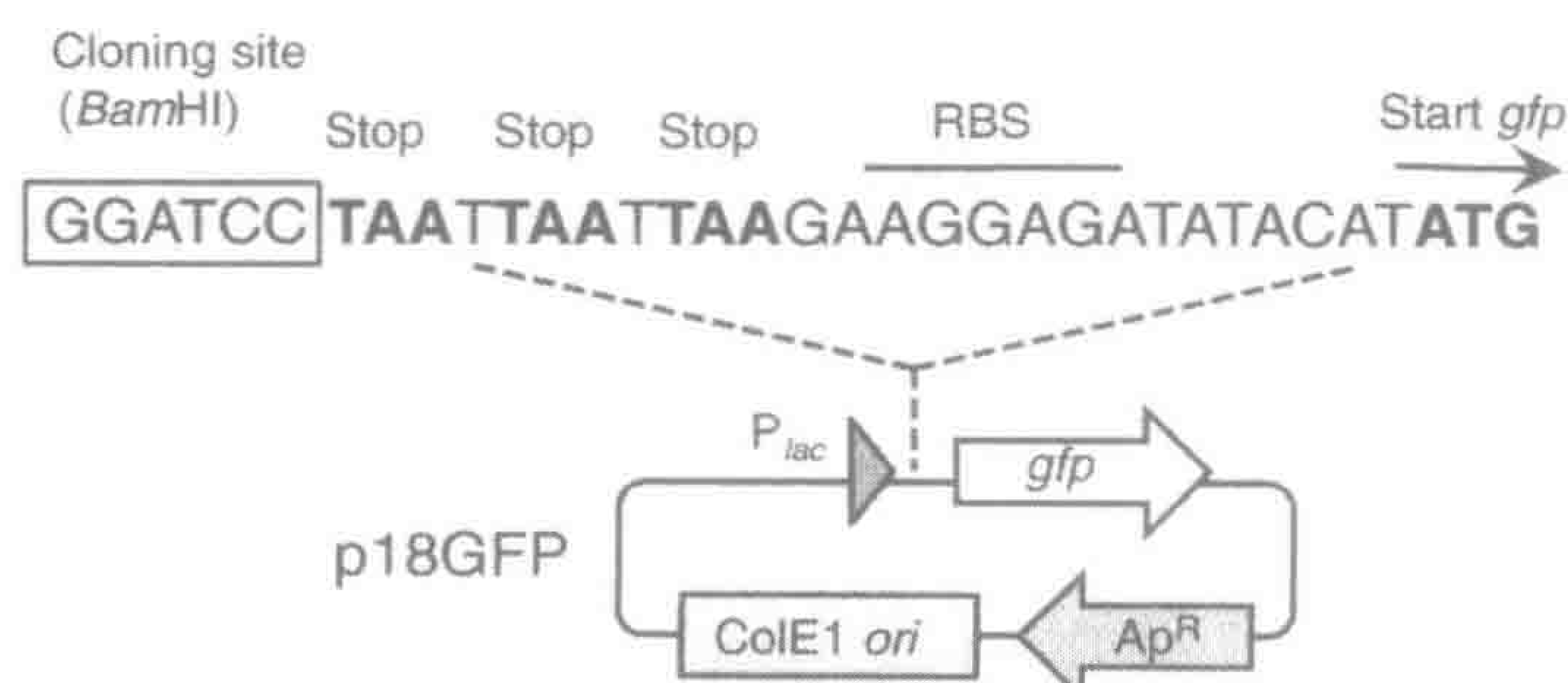


Fig. 2. The operon-trap vector p18GFP. The vector includes the *gfp* gene encoding green fluorescent protein (GFP), which is used as a coexpressed reporter. GFP fluorescence is used to detect clones that contain metagenome fragments with substrate-induced gene-expression activity. The *Bam*HI cloning site precedes stop codons in three frames, a ribosome-binding site (RBS), and the *gfp* start codon. It also includes the *lac* promoter (P_{lac}), which is used to detect and eliminate clones harboring self-ligated p18GFP.

Step 3 is to select clones that fluoresce in response to substrate. Cells selected in Step 2 are grown in the presence of substrate and subjected to FACS to select fluorescent cells. A wide variety of chemicals can be used as induction substrates, with the caveat that the induction substrate must be able to permeate into the cytoplasm of the cloning host. Substrates of the target catabolic enzymes are commonly used as inducers. However, to promote the discovery of novel biological events in addition to the targeted genes, we recommend using as inducers a series of related compounds, such as substrate analogues, possible metabolites, and products, in addition to the substrates.

In Step 4, the positive clones are isolated. Because the sorted cells will undoubtedly contain multiple copies of some clones, it is essential to identify the independent clones in the sorted cells. After fluorescent cells are spread on agar plates to obtain single colonies, several colonies are picked, and their plasmids are purified and analyzed for restriction fragment length polymorphisms (RFLPs). The results allow identification of independent clones, which are, subsequently, further characterized by methods such as DNA sequencing and enzymatic activity assays.

DNA fragments obtained by SIGEX are expected to contain catabolic operons that are related to the induction substrates. However, sometimes the retrieved gene fragment does not contain a full-length copy of the gene of interest, in which case the gene-walking technique must be performed to clone the flanking regions. Although gene walking is an established method, it is rather difficult to perform using the metagenome as a template. It is thus advisable to employ a gene-walking technique such as IAN-PCR (9), which was specifically developed for difficult metagenome-walking.

3.1. Construction of an Operon-Trap Vector

1. Construct or select an appropriate operon-trap vector (e.g., p18GFP; Fig. 2) by considering the cloning host and screening system to be used.

3.2. Verify Construction of the Operon-Trap Vector System

1. Transform *E. coli* JM109 competent cells with p18GFP using electroporation (10). Select transformants on LB agar plates containing ampicillin (LB-Ap plates). Incubate at 37°C overnight.
2. Pick single colonies from the plates and inoculate them into separate liquid LB-Ap cultures. Incubate at 30°C with vigorous shaking (200 cycles/min) until the optical density at 600 nm (OD_{600}) reaches 0.6.
3. Induce GFP expression by adding IPTG (0.5 mM). Incubate at 30°C overnight with vigorous shaking (200 cycles/min).

4. Centrifuge the cultures and remove the supernatant fraction. Wash the cell pellets twice with PBS, resuspend the cells in fresh PBS, and immediately use them to check for GFP expression (see Note 3).

3.3. Verification of GFP Expression by Flow Cytometry

1. Dilute the cell suspension with PBS to an appropriate concentration (10^5 to 10^7 cells/mL) (see Note 4).
2. If using a FACS Vantage SE flow cytometer, input the following settings: laser output power = 0.5 W; nozzle tip diameter = 70 μm; sheath pressure = 11 psi. (For other flow cytometers, these settings will need to be modified. These settings are optimized for the FACS Vantage SE.) Fluorescent calibration beads are used for the laser alignment setting.
3. Check the flow cytometer parameters by comparing the fluorescence properties of non-GFP-expressing cells (negative-control sample; e.g., cells transformed with pUC18) and GFP-expressing cells (positive-control sample; e.g., cells transformed with p18GFP). First, determine flow cytometer parameters [forward scatter (FSC), side scatter (SSC), and green fluorescence (FL1)] by checking the light scattering and fluorescence properties of the negative-control sample (Fig. 3a). Initial instrument settings [threshold parameter, photomultiplier (PMT) voltage, and detector sensitivity settings] should be set as follows:

   Threshold parameter: SSC, value 50

   FSC sensitivity: logarithmic amplification

   SSC sensitivity: 280 V, logarithmic amplification

   FL1 sensitivity: 600 V, logarithmic amplification
4. Place the negative-control sample in a sample tube and determine the event rate. A rate of less than 3,000 events/s is necessary for high-resolution analysis. If necessary, the event rate can be decreased by reducing the flow rate or by diluting the sample.
5. Optimize the instrument settings. Adjust photomultiplier tube (PMT) voltages and threshold levels. Check cell count vs. log FSC and cell count vs. log SSC histograms (upper two panels in Fig. 3a) to verify that the edges of the bell-shaped peak are not cut off (see Note 5).
6. Generate a dot plot of the SSC vs. FSC. Then set a gate R1 around the bacterial population (lower left panel in Fig. 3a). Statistical information for R1 is provided in the CellQuest Pro program (BD Biosciences) (see Note 6).
7. For GFP fluorescence intensity measurements, the FL1 PMT voltage should be adjusted using the negative-control sample. Use a histogram (log FL1 fluorescence vs. cell count) to ana-

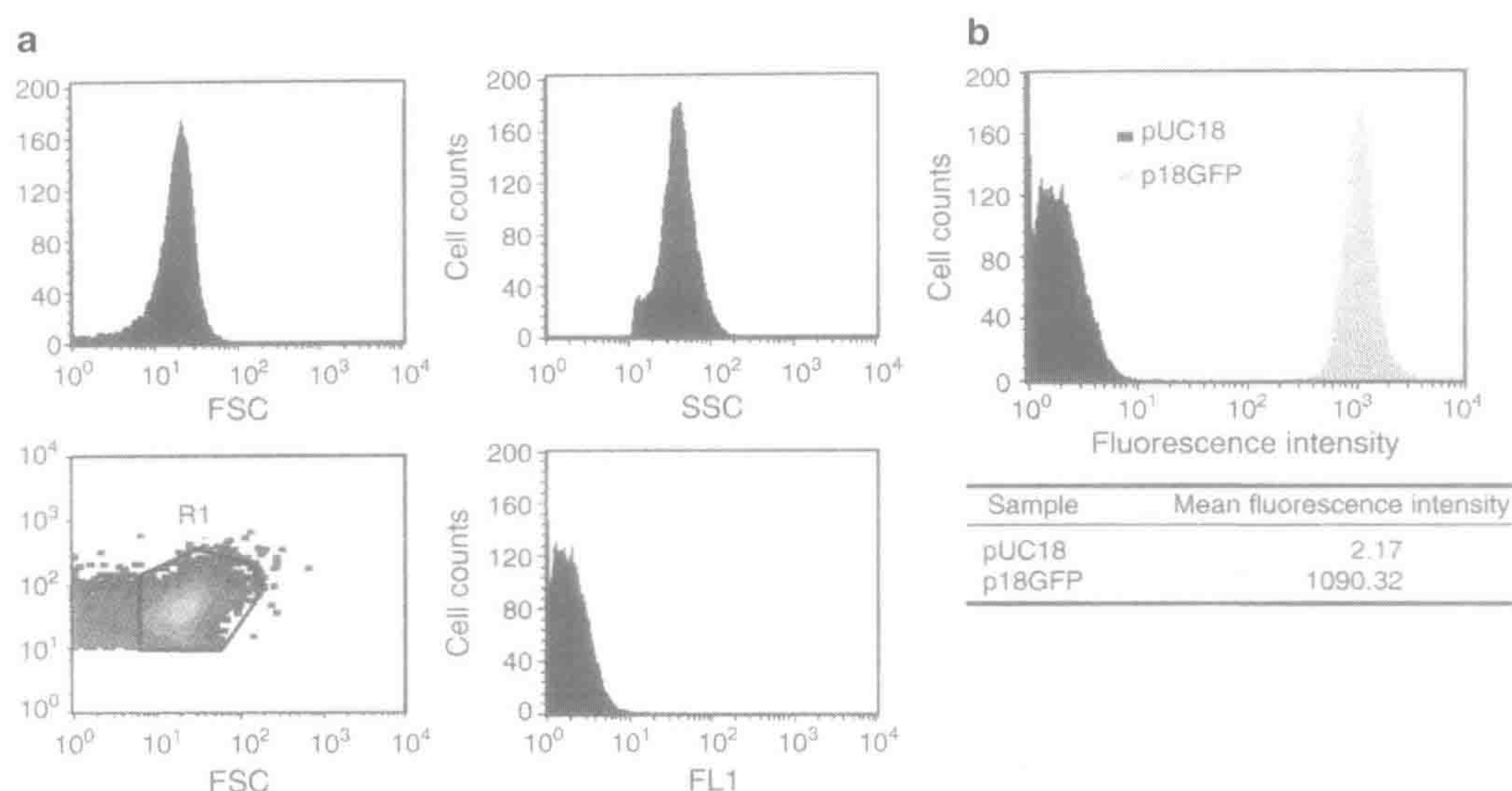


Fig. 3. Gating in FACS analysis and comparison of GFP fluorescence in positive- and negative-control samples. (**a**) Cells were analyzed in forward scatter (FSC) and side scatter (SSC) histograms (*upper two panels*), and also in FSC vs. SSC dot plots (*lower left panel*). To select an appropriate particle size (that of bacterial cells), an appropriate gate (R1) was set and the fluorescence intensities (FL1) of particles in the gate were analyzed (*lower right panel*). (**b**) Comparative FL1 histogram of positive- and negative-control samples. The histograms obtained from the negative-control sample [*E. coli* transformed with pUC18 (*black*; mean fluorescence intensity = 2.17)] and positive-control sample [*E. coli* transformed with p18GFP (*gray*; mean fluorescence intensity = 1,090.32)] are superimposed.

lyze the fluorescence of the negative-control sample (lower right panel in Fig. 3a). Set the PMT voltage so that the mean fluorescence value of the peak is less than 3 (the entire peak area should be less than 10). Acquire a total of 20,000 events for R1. Flow cytometer settings used for detecting the negative control sample (*E. coli* JM109 transformed with pUC18) in this study are as follows:

Threshold parameter: SSC, value 180

SSC sensitivity: 310 V, logarithmic amplification

FL1 sensitivity: 600 V, logarithmic amplification

8. After the measurements, detach the sample tube from the instrument. Wash the sample lines by flushing the sheath fluid to remove residual bacterial cells.

9. Analyze the positive-control sample (e.g., *E. coli* JM109 transformed with p18GFP) using the settings established for the negative-control sample. Acquire a total of 20,000 events for R1. Statistical information for R1 is provided in the CellQuest Pro program.

10. Compare the histograms. Examine the mean fluorescence intensities of the positive- and negative-control samples (Fig. 3b). The mean fluorescence intensity value correlates with the level of gene-induction activity encoded in the clone fragment.

3.4. Preparation of Metagenomic DNA

1. Place 5 g of soil into a plastic tube. Add 13.5 mL of DNA-extraction buffer and 0.1 mL of proteinase K solution to the tube, and suspend the soil by vortexing. Incubate the tube at 37°C for 30 min with horizontal shaking on a reciprocal shaker at 200 cycles/min.
2. Add 1.5 mL of 20% SDS and 750 µL of lysozyme solution to the tube. Incubate at 65°C for 3 h with horizontal rocking on a reciprocal shaker at 60 cycles/min. After rocking, centrifuge the suspension at 6,000 × *g* for 20 min. Transfer the supernatant fraction into a fresh tube using a wide-bore pipette tip.
3. Add an equal volume of TE-saturated phenol to the supernatant fraction, and incubate with horizontal rocking (60 cycles/min) overnight.
4. Centrifuge the suspension at 6,000 × *g* for 20 min. Transfer the aqueous phase into a fresh tube using a wide-bore pipette tip.
5. Add an equal volume of TE-saturated phenol to the aqueous phase and incubate with horizontal rocking (60 cycles/min) for 1 h. Centrifuge the suspension at 6,000 × *g* for 20 min. Transfer the aqueous phase into a fresh tube using a wide-bore pipette tip.
6. Add an equal volume of PCI, and incubate with horizontal rocking (60 cycles/min) for 1 h. Centrifuge the suspension at 6,000 × *g* for 20 min. Transfer the aqueous phase into a fresh tube using a wide-bore pipette tip.
7. To the tube containing the aqueous phase, add 0.1 volume of 3 M sodium acetate (pH 5.2) and 2 volumes of ethanol. Mix gently until the ethanol solution is thoroughly mixed. The DNA will immediately form a precipitate. Remove the DNA precipitate from the ethanol solution with a Pasteur pipette that has been modified to have a sealed "J"-shaped tip.
8. Wash the DNA precipitate two times with 70% ethanol and collect the DNA by centrifugation at 6,000 × *g* for 5 min. Store the DNA pellet in an open tube until the ethanol has evaporated (see Note 7).
9. Add 1 mL of TE buffer to the dry DNA pellet. Incubate the solution with rocking (60 cycles/min) at room temperature overnight.
10. The following day, analyze the quality of the metagenomic DNA preparation by electrophoresing it through an agarose gel (11,12). Good-quality metagenomic DNA consists of DNA fragments larger than 20 kb. Large DNA fragments (>20 kb) can be used in the subsequent steps to construct a metagenomic library (see Note 8).

3.5. Construction of a Metagenomic Library

1. Linearize the vector DNA: Digest 5 μg of the operon-trap vector DNA (e.g., p18GFP) with the restriction enzyme *Bam*HI using twice the amount of *Bam*HI recommended by the manufacturer. Incubate the digestion reaction at 30°C for 5 h. Afterward, recover the DNA using the ethanol precipitation procedure (13) and dissolve it in TE buffer.
2. Dephosphorylate the DNA: To 1 μg of the linearized vector DNA, add an appropriate amount (as recommended by the manufacturer) of shrimp alkaline phosphatase and incubate the mixture at 37°C for 1 h. Then, incubate the reaction mixture for 15 min at 65°C to inactivate the enzyme.
3. Recover the DNA using the ethanol precipitation procedure (13) and dissolve the linearized and dephosphorylated vector DNA in TE buffer.
4. Partially digest the metagenomic DNA with the restriction enzyme *Sau*3AI. Conditions for the partial digestion of the DNA into fragments of the desired length should be optimized in a small-scale pilot experiment. (In our experiments, fragments of 5–10 kb were used in construction of the metagenomic library.)
5. After the partial digestion, separate the DNA fragments by agarose gel electrophoresis (11,12). A good-quality sample will exhibit a high concentration of fragments of the desired molecular sizes (see Note 9).
6. Using a scalpel, make an incision in the agarose gel in front of (on the + side) the DNA bands. (In our experiments, a range of 5–10 kb was recovered.) Into the incision, insert a RECOCHIP previously treated with electrophoresis buffer.
7. Replace the gel in the electrophoresis apparatus and run the apparatus for a few minutes. Stop the electrophoresis and remove the RECOCHIP from the gel.
8. Put the RECOCHIP into a 2-mL tube and centrifuge it at 4,020 × *g* for 5 s. Recover the DNA solution, purify the DNA using the ethanolprecipitation procedure (13) and dissolve the DNA in TE buffer.
9. Prepare a ligation mixture using the following components:
   (a) Vector DNA [linearized (by *Bam*HI) and dephosphorylated p18GFP]: 50 ng
   (b) Insert DNA (5- to 10-kb fragments of partially *Sau*3AI-digested metagenomic DNA): 25 ng
   (c) 10× LONG ligation buffer (from DNA ligation kit LONG): 5 μL
   (d) H_2O: To bring volume to 49 μL

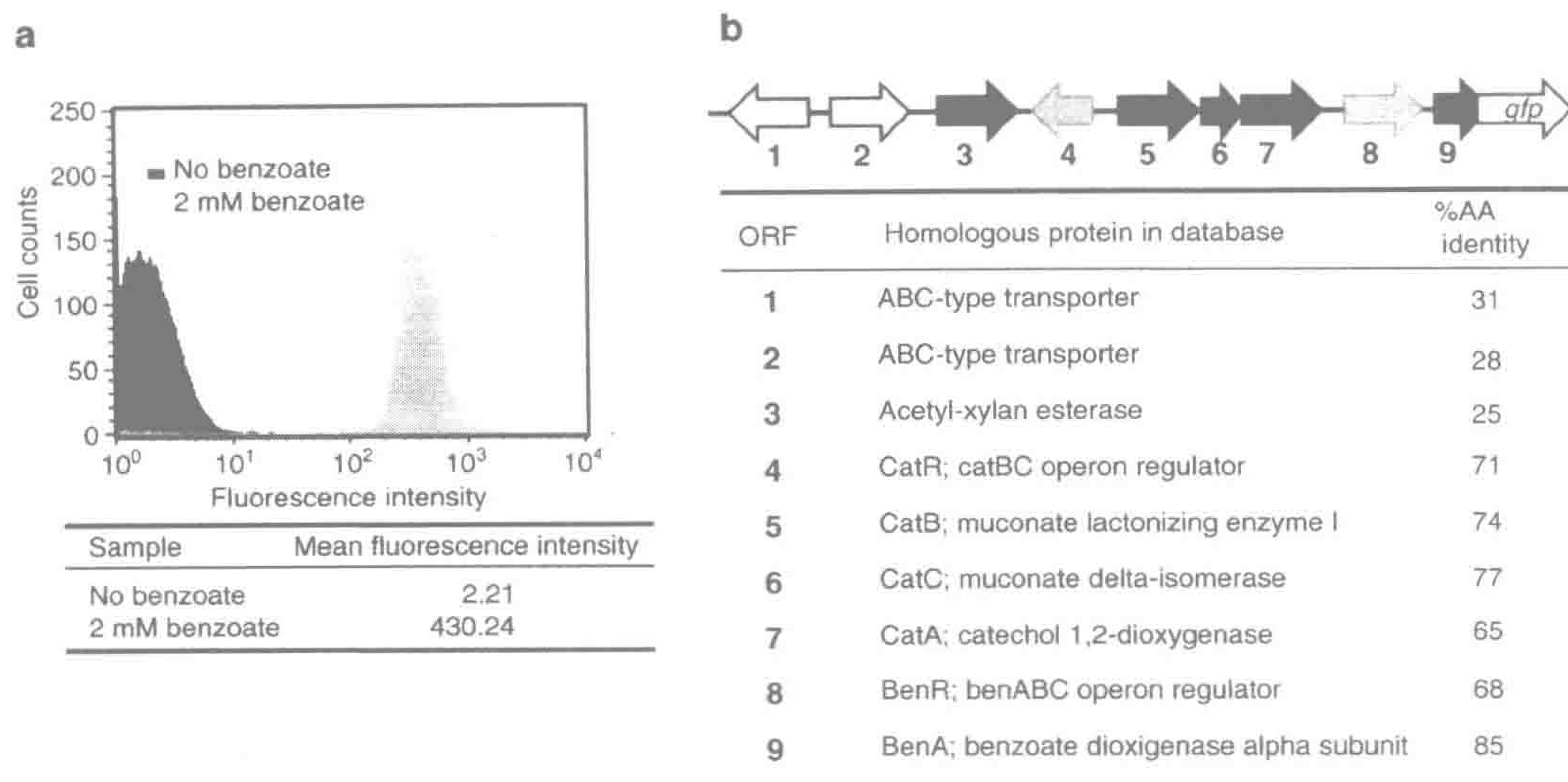
10. Incubate the mixture at 65°C for 3 min and cool immediately on ice. Add 1 μL of DNA ligase (from DNA ligation kit LONG) to the reaction mixture. Incubate at 16°C overnight.
11. Recover the DNA using the standard ethanol precipitation procedure (13), and dissolve it in TE buffer.
12. Transform competent cells (*E. coli* JM109) with the ligation mixture using electroporation (10) (see Note 10). After the transformation, inoculate the cells into a large culture tube containing 1 mL of SOC medium and incubate at 37°C for 1 h with gentle rotation.
13. Add 4 mL of LB-Ap to the culture tube and incubate at 37°C overnight with vigorous shaking (200 cycles/min) on a rotary shaker. This culture constitutes an original metagenomic library. It can be stored at −80°C without significant loss of viability if the medium is supplemented with 15% (wt/vol) glycerol.
14. Take an aliquot (~10 μL) from the culture and spread it on an LB-Ap plate containing IPTG (0.5 mM). Incubate the plate at 37°C overnight.
15. Place the plate on a UV transilluminator (365 nm). Count non-GFP-expressing (white) colonies on the plate (see Note 11). Pick ten white colonies and inoculate them individually into culture tubes containing 1 mL of LB-Ap. Incubate at 37°C overnight with vigorous shaking (200 cycles/min) on a rotary shaker.
16. Extract plasmid DNA from the cultures using a standard plasmid extraction procedure (14). Analyze the size of the inserted DNA fragments by restriction enzyme digestion and agarose gel electrophoresis (11,12). A good-quality library has random inserts.

3.6. FACS Screening

1. Inoculate 30 μL of the first metagenomic library into a culture tube containing 3 mL of LB-Ap. Incubate at 30°C with vigorous shaking (200 cycles/min) on a rotary shaker until OD_{600} reaches 0.6.
2. Induce GFP expression by adding IPTG (0.5 mM). Incubate at 30°C overnight with vigorous shaking (200 cycles/min).
3. The following day, set up the flow cytometer and prepare the cell sorter. Apply the cells from the first metagenomic library to the flow cytometer. Set a sorting region for cells with no fluorescence signal. In this step, fluorescent cells contain either the self-ligated vector or vectors containing constitutively expressed inserts.

4. Check the efficiency of the sorting by resorting the sorted sample. In the sorted fraction, the ratio of nonfluorescent cells to total cells should be increased significantly.
5. Inoculate the sorted cells into a culture tube containing 5 mL of LB-Ap. Incubate at 37°C overnight with vigorous shaking (200 cycles/min). The following day, determine the ratio of nonfluorescent cells to total cells in the culture using the flow cytometer. If the nonfluorescent cells represent a high percentage (98%) of the total cell population, then the cells containing the self-ligated vector or vectors ligated with constitutively expressed inserts have been successfully removed from the library. This culture constitutes the ***second metagenomic library***. If the nonfluorescent cells represent less than 98% of the total cell population, repeat steps 1–5.
6. Inoculate an aliquot (10 μL) of the second metagenomic library culture into a culture tube containing 1 mL of dLB medium containing ampicillin (100 μg/mL) (dLB-Ap). Incubate at 30°C with vigorous shaking (200 cycles/min) until OD_{600} reaches 0.4.
7. Induce GFP expression by adding an induction substrate (e.g., benzoate in our previous study) at a final concentration of 2 mM. Incubate the culture at 30°C overnight with vigorous shaking (200 cycles/min).
8. The following day, set up the flow cytometer and prepare the cell sorter. Apply the culture to the flow cytometer. In the resulting histogram (cell count vs. log FL1), set a sorting region to collect cells expressing GFP. In our study, this gating sorted out ~500 cells.
9. Plate the sorted cells onto LB-Ap plates containing IPTG (0.5 mM) and incubate the plates at 37°C overnight.
10. Lay the plates on a UV transilluminator (365 nm), pick white colonies, and inoculate them into culture tubes containing 1 mL of dLB-Ap. Incubate the tubes at 30°C with vigorous shaking (200 cycles/min) until OD_{600} reaches 0.4.
11. Divide the cultures into two fractions. To one of the fractions, add the induction substrate at a final concentration of 2 mM. Incubate both fractions at 30°C overnight with vigorous shaking (200 cycles/min).
12. The following day, compare the fluorescence intensities of the induced and noninduced cultures using the flow cytometer (Fig. 4a). Determine the induction efficiency using the equation below:

    Induction efficiency = (mean fluorescence intensity of the induced culture)/(mean fluorescence intensity of the noninduced culture).



Sample	Mean fluorescence intensity
No benzoate	2.21
2 mM benzoate	430.24

ORF	Homologous protein in database	%AA identity
1	ABC-type transporter	31
2	ABC-type transporter	28
3	Acetyl-xylan esterase	25
4	CatR; catBC operon regulator	71
5	CatB; muconate lactonizing enzyme I	74
6	CatC; muconate delta-isomerase	77
7	CatA; catechol 1,2-dioxygenase	65
8	BenR; benABC operon regulator	68
9	BenA; benzoate dioxigenase alpha subunit	85

Fig. 4. Verification of GFP induction in and sequence analysis of the benzoate-responsive clone *pbzo26* (6). (**a**) FL1 histogram of *pbzo26*. The histograms obtained under noninduced conditions (*black*) and after induction with 2 mM benzoate (*gray*) are superimposed. The mean FL1 intensity under noninduced conditions was 2.21, whereas it was 430.24 after induction with benzoate. The induction efficiency was 194.68. (**b**) Schematic representation of *pbzo26*. The clone contains open-reading frames (ORFs) homologous to genes in catechol-degradative (16) and benzoate-degradative (17) operons.

13. Extract plasmid DNA from a positive clone using a standard plasmid extraction procedure (14), and analyze the nucleotide sequence (15) of the inserted DNA fragment (Fig. 4b) (see Note 12).

4. Notes

1. TE-saturated phenol is toxic. Wear gloves to avoid direct contact with the solution, and use in a chemical fume hood.
2. PCI is toxic. Wear gloves to avoid direct contact with the solution, and use in a chemical fume hood.
3. Washing is essential to completely remove the medium because trace contamination of medium causes background fluorescence.
4. Dilution prevents aggregation of the cells and prevents clogging of the narrow-bore tube of the flow cytometer. Alternatively, pass the cell suspension through a nylon-mesh filter to remove cell aggregates.
5. Sometimes, many peaks are observed in the cell count vs. log FSC histograms. This phenomenon can occur when the

library host (e.g., *E. coli*) undergoes changes in cell size, cell shape, or aggregation trend. These changes can be identified by microscopy.

6. Use logarithmic amplification settings for FSC, SSC, and FL1 detector sensitivity so that wide ranges of bacterial sizes and fluorescence intensities can be observed, which facilitates gate setting. However, the small size of bacteria can lead to problems in the establishment of a recognizable population in the SSC vs. FSC dot plot. Analyze a blank sample (cell-suspension buffer) to check the background noise level. Filtering and autoclaving of the cell-suspension buffer and sheath fluid used to suspend the bacteria should minimize background particle-scatter signal noise.
7. Do not allow the DNA pellet to dry completely. Completely dry DNA is very difficult to dissolve.
8. Metagenomic DNA samples should be stored for further analysis, which may include post-SIGEX analysis. For instance, when only a partial gene is obtained by SIGEX, further work is needed to obtain the flanking genome fragment [e.g., using IAN-PCR (9)].
9. Occasionally, restriction enzyme digestion of DNA extracted from an environmental sample is problematic, possibly because the DNA can be contaminated with proteins, mucoids, and humic acids. Further purification of the DNA (e.g., through the use of a commercial column-chromatography kit) may resolve the problem.
10. We recommend electroporation (10) as a method for achieving the highest transformation efficiency. A negative-control reaction (operon-trap vector alone) must be used to estimate the number of background transformants caused by self-ligation of the vector.
11. When using competent cells with a transformation efficiency of 1×10^8 cfu/μg, the above-described ligation and electroporation procedures yield more than 1×10^4 non-GFP-expressing (white) colonies. GFP-expressing (green) colonies on the plates contain either the self-ligated vector or vectors containing constitutively expressed inserts.
12. RFLP analysis is useful for verifying independence among the clones.

Acknowledgements

This work was supported by the Japan Society for Promotion of Science (No. 19880040).

References

1. Amann, R. I., Ludwig, W., and Schleifer, K. H. (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* 59, 143–169.
2. Cowan, D. A. (2000) Microbial genomes – the untapped resource. *Trends Biotechnol* 18, 14–16.
3. Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., and Goodman, R. M. (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5, 245–249.
4. Lorenz, P., Liebeton, K., Niehaus, F., and Eck, J. (2002) Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. *Curr Opin Biotechnol* 13, 572–577.
5. Ferrer, M., Martínez-Abarca, F., and Golyshin, P. N. (2005) Mining genomes and 'metagenomes' for novel catalysts. *Curr Opin Biotechnol* 16, 1–6.
6. Uchiyama, T., Abe, T., Ikemura, T., and Watanabe, K. (2005) Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat Biotechnol* 23, 88–93.
7. Uchiyama, T. and Watanabe, K. (2007) The SIGEX scheme: high throughput screening of environmental metagenomes for the isolation of novel catabolic genes. *Biotechnol Genet Eng Rev* 24, 107–116.
8. Uchiyama, T. and Watanabe, K. (2008) Substrate-induced gene expression (SIGEX) screening of metagenome libraries. *Nat Protoc* 3, 1202–1212.
9. Uchiyama, T. and Watanabe, K. (2006) Improved inverse PCR scheme for metagenome walking. *BioTechniques* 41, 183–188.
10. Sambrook, J. and Russell, D. W. (2001) Transformation of *E. coli* by electroporation. In: Sambrook, J. and Russell, D. W. Ed. Molecular Cloning: A Laboratory Manual, 3rd Edn., Vol. 1. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 1.119–1.122.
11. Sambrook, J. and Russell, D. W. (2001) Agarose gel electrophoresis. In: Sambrook, J. and Russell, D. W. Ed. Molecular Cloning: A Laboratory Manual, 3rd Edn., Vol. 1. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 5.4–5.13.
12. Sambrook, J. and Russell, D. W. (2001) Detection of DNA in agarose gels. In: Sambrook, J. and Russell, D. W. Ed. Molecular Cloning: A Laboratory Manual, 3rd Edn., Vol. 1. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 5.14–5.17.
13. Sambrook, J. and Russell, D. W. (2001) Commonly used techniques in molecular cloning. In: Sambrook, J. and Russell, D. W. Ed. Molecular Cloning: A Laboratory Manual, 3rd Edn., Vol. 3. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. A8.12–A8.16.
14. Sambrook, J. and Russell, D. W. (2001) Preparation of plasmid DNA by alkaline lysis with SDS. In: Sambrook, J. and Russell, D. W. Ed. Molecular Cloning: A Laboratory Manual, 3rd Edn., Vol. 1. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 1.31–1.42.
15. Sambrook, J. and Russell, D. W. (2001) DNA sequencing. In: Sambrook, J. and Russell, D. W. Ed. Molecular Cloning: A Laboratory Manual, 3rd Edn., Vol. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 12.1–12.120.
16. Rothmel, R. K., Aldrich, T. L., Houghton, J. E., Coco, W. M., Ornston, L. N., and Charkrabaty, A. M. (1990) Nucleotide sequencing and characterization of *Pseudomonas putida catR*: a positive regulator of the *catBC* operon is a member of the LysR family. *J Bacteriol* 172, 922–931.
17. Cowles, C. E., Nichols, N. N., and Harwood, C. S. (2000) BenR, a XylS homolog, regulates three different pathways of aromatic acid degradation in *Pseudomonas putida*. *J Bacteriol* 182, 6339–6346.

Chapter 11

Screens for Active and Stereoselective Hydrolytic Enzymes

Dominique Böttcher, Marlen Schmidt, and Uwe T. Bornscheuer

Abstract

A procedure for the high-throughput screening (HTS) of esterases is described. This includes a pretest for discrimination of active and inactive clones using an agar plate overlay assay, the enzyme expression in microtiter plates and the measurement of activity and enantioselectivity (E) of the esterase variants using acetates of secondary alcohols as model substrates. Acetic acid released is converted in an enzyme cascade leading to the stoichiometric formation of NADH, which is quantified in a spectrophotometer. The method allows screening of several thousand mutants per day and has already been successfully applied to identify an esterase mutant with an $E > 100$ towards an important building block for organic synthesis. This protocol can also be used for lipases and possibly other hydrolases that are expressed in soluble form in conventional *Escherichia coli* strains.

Key words: Hydrolase, Esterase, Lipase, High-throughput assay, Enantioselectivity, Directed evolution, Metagenome

1. Introduction

Lipases and esterases are the most frequently used hydrolases (EC 3) in organic synthesis (1). They are important biocatalysts and especially suitable for industrial applications as they are very stable and also active in organic solvents. Moreover, they very often exhibit high enantioselectivity and are therefore used in the synthesis of optically active compounds, for which more than 1,000 examples can be found in literature. Besides a considerable number of commercially available lipases and to a lesser extent esterases, researchers can create optimized enzyme variants using directed evolution (2–4) experiments or identify new esterases or lipases with desired activity/selectivity using the metagenomic approach (5–7).

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668,
DOI 10.1007/978-1-60761-823-2_11,

These methods can create huge numbers of novel biocatalysts, which are time-consuming to screen using conventional methods such as gas chromatography or HPLC.

Consequently, a range of high-throughput assay systems has been developed in the past few years to allow for a rapid and reliable identification of suitable enzymes (8–11). As lipases and esterases are often used to produce optically active compounds, the determination of the enantioselectivity of the enzymes is of major interest, and several methods have been described (12) and successfully applied in directed evolution experiments to improve the biocatalysts' selectivity (4, 13, 14).

The protocol described here was designed to allow the determination of substrate specificity and enantioselectivity of esterases (or other hydrolases such as peptidases and amidases, where an acetamide can serve as substrate) towards secondary alcohols and has the advantage that no surrogate substrates (i.e., chromophores like resorufin) have to be used, as acetates are the preferred esters in the resolution of alcohols. In this assay, hydrolysis of the acetate using an esterase (or lipase) releases acetic acid. This is then converted in an enzyme cascade reaction into citrate with stoichiometrical formation of NADH. This increase in NADH concentration is quantified spectrophotometrically at 340 nm (15) (Fig. 1). This HTS is very reliable and fast, as the exact determination of activity and enantioselectivity is possible within minutes for an entire 96-well plate. In addition, the acetic acid kit is commercially available (R-Biopharm GmbH, Darmstadt, Germany).

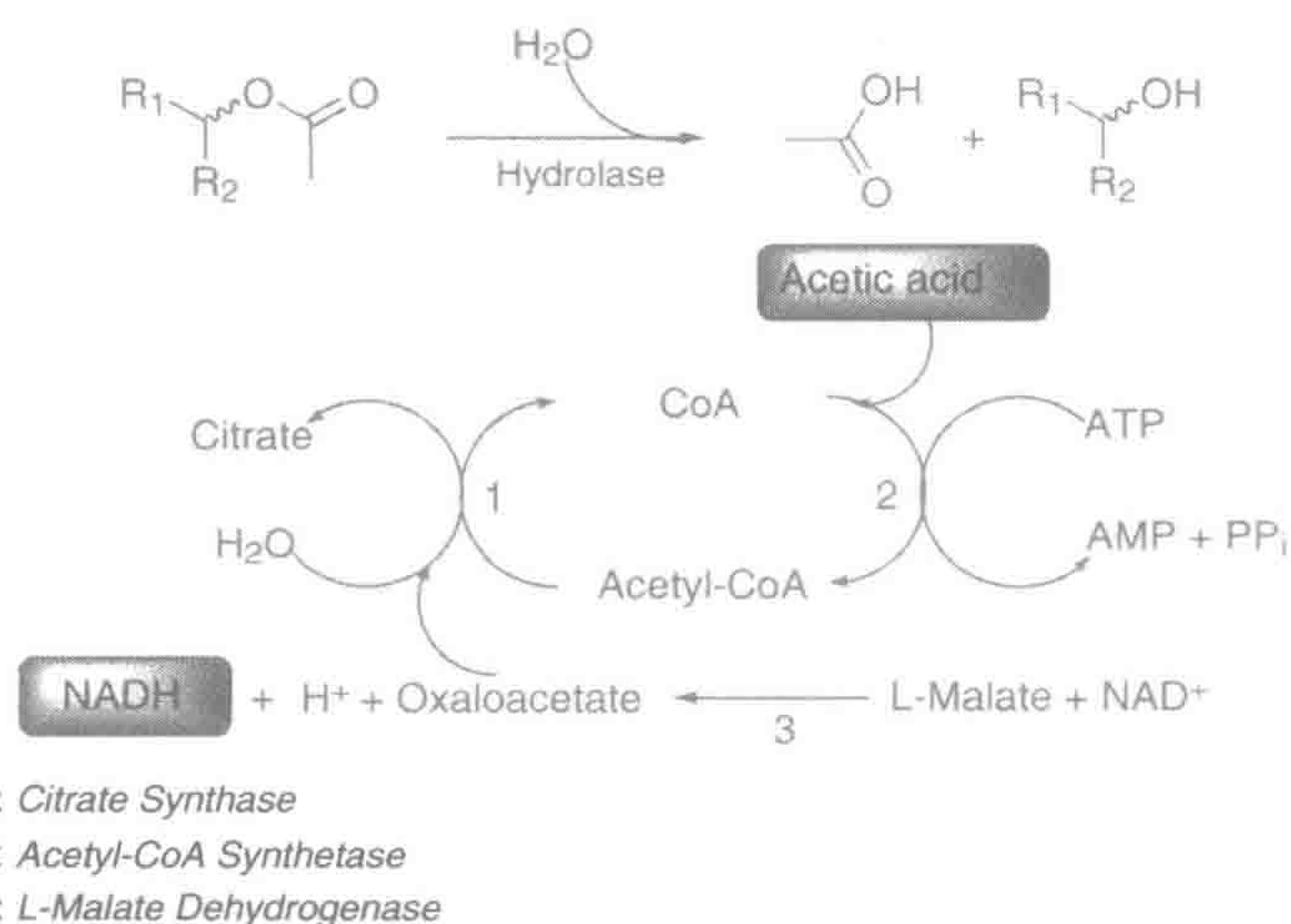


Fig. 1. Assay based on the conversion of acetic acid released in the hydrolase-catalyzed reaction in a subsequent enzyme cascade yielding an increase in NADH (15).

2. Materials

2.1. Activity Test on Agar Plates (Overlay Agar)

1. Agar plates containing colonies from metagenomic libraries.
2. Replica-plating tool and sterile clothes.
3. Soft agar (0.5% agar dissolved in water).
4. 1-Naphthyl acetate solution: 40 mg/mL in *N*,*N'*-dimethyl formamide.
5. Fast Red TR solution: 100 mg/mL in dimethyl sulfoxide.

2.2. Cultivation in Microtiter Plates

1. Luria–Bertani (LB) medium: 10 g tryptone/peptone, 10 g NaCl, 5 g yeast extract, add H_2O to 1,000 mL.
2. 60% Glycerol.
3. Antibiotic (e.g., ampicillin, usually 100 μg/mL).
4. Isopropyl-β-D-thiogalactoside (IPTG).
5. 96-Well microtiter plates (e.g., Greiner Bio-One, Kremsmunster, Austria).
6. Microtiter plate thermoshaker (e.g., Biosan, Riga, Latvia), alternatively use a plastic box with wet tissue placed in a normal incubator.
7. Centrifuge with microtiter plate adapter (e.g., Heraeus Labofuge 400R, Thermo Electron Corporation, Waltham, MA, USA).

2.3. Cell Lysis

1. Lysis buffer: 50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0, 0.1% (w/v) lysozyme, 1 U/mL DNaseI.

2.4. Enzyme Assay

1. 10 mM Sodium phosphate buffer, pH 7.4.
2. Acetate assay reagents (R-Biopharm GmbH, Darmstadt, Germany).
   (a) Bottle 1: 32 mL triethanolamine buffer solution pH 8.4, L-malic acid (134 mg), $MgCl_2 \cdot 6H_2O$ (67 mg), storage at 2–8°C.
   (b) Bottle 2: lyophilizate containing ATP (175 mg), CoA (18 mg), NAD^+ (86 mg) dissolve in 7 mL distilled water, aliquots are stable at –20°C for 2 months.
   (c) Bottle 3: suspension of L-malate dehydrogenase (1,100 U); citrate synthase (270 U); stable at 2–8°C.
   (d) Bottle 4: lyophilized acetyl-CoA synthetase (5 U) add 250 μL distilled water, stable at 2–8°C for 5 days.

   To prepare the test-kit component mixture use 1,000 μL of bottle 1, 200 μL of bottle 2, 10 μL of bottle 3, 20 μL of bottle 4, and 1,900 μL distilled water.

3. Racemic (A) or enantiopure (B) acetate substrates 5–50 mM.
4. Multichannel pipette.
5. Microtiter plate spectrophotometer (e.g., FLUOstar Galaxy, BMG, Offenburg, Germany).

Optional:

1. Colony picking robot (e.g., Biopick, BioRobotics, Inc., Woburn, MA, USA).
2. Pipetting robot (Tecan MiniPrep75, Tecan, Mannedorf/Zurich, Switzerland).
3. 96-Pin replicator (Thermo Fisher Scientific, Waltham, MA, USA).

3. Methods

The acetic acid assay allows to differentiate active from inactive and enantioselective from nonselective enzyme variants. In the activity test (option A), the resulting graphs will provide only the relative activity of each enzyme variant for the tested acetate ester. In the selectivity test (option B), one has to calculate the initial reaction rates ($\Delta A/\Delta t$) for each enantiomer separately, and the quotient of the two rates then yields the apparent enantioselectivity E_{app}.

Positive hits from the assay must be verified afterwards using conventional analytical methods, such as chiral GC or HPLC to determine conversion, kinetic parameters, and the true enantioselectivity E_{true}.

3.1. Activity Test on Agar Plates

1. Spread cells containing metagenomic library onto LB agar plates containing an appropriate antibiotic.
2. Incubate the plates overnight at 30 or 37°C.
3. Transfer colonies by replica plating to LB agar plates containing an appropriate antibiotic and IPTG to induce esterase production.
4. Incubate the plates for 5 h at 37°C.
5. Prepare overlay soft agar.
6. Prepare solutions of 1-naphthyl acetate and Fast Red TR.
7. Melt the soft agar in a microwave and let it cool down to approximately 40°C.
8. Mix 100 μL of both solutions with 10 mL soft agar and pour it carefully over the colonies.
9. Active clones will turn brownish in few seconds.

3.2. Enzyme Production in Microtiter Plates (Fig. 2)

1. Pick single colonies into 96-well microtiter plates containing 200 μL LB medium, supplemented with the required antibiotic, per well. These plates serve as master plates. After cell growth for 24 h at 37°C and 220 rpm, duplicate the master plates by transferring a 2 μL aliquot (see Note 1) to a new microtiter plate (containing 200 μL LB-antibiotic medium per well) used for the subsequent production of esterase (production plates).
2. Supplement the master plates with glycerol (final concentration 15% v/v) and store them at –80°C. These master plates can be also used for future high-throughput assays.
3. Incubate the production plates overnight at 37°C and 220 rpm and dilute 1:10 the next day with fresh medium (see Note 2). Cultivate at 37°C and 220 rpm.
4. After 4 h start enzyme production by the addition of inducer solution in an appropriate concentration (e.g., IPTG usually in concentrations from 10 to 1,000 μM).

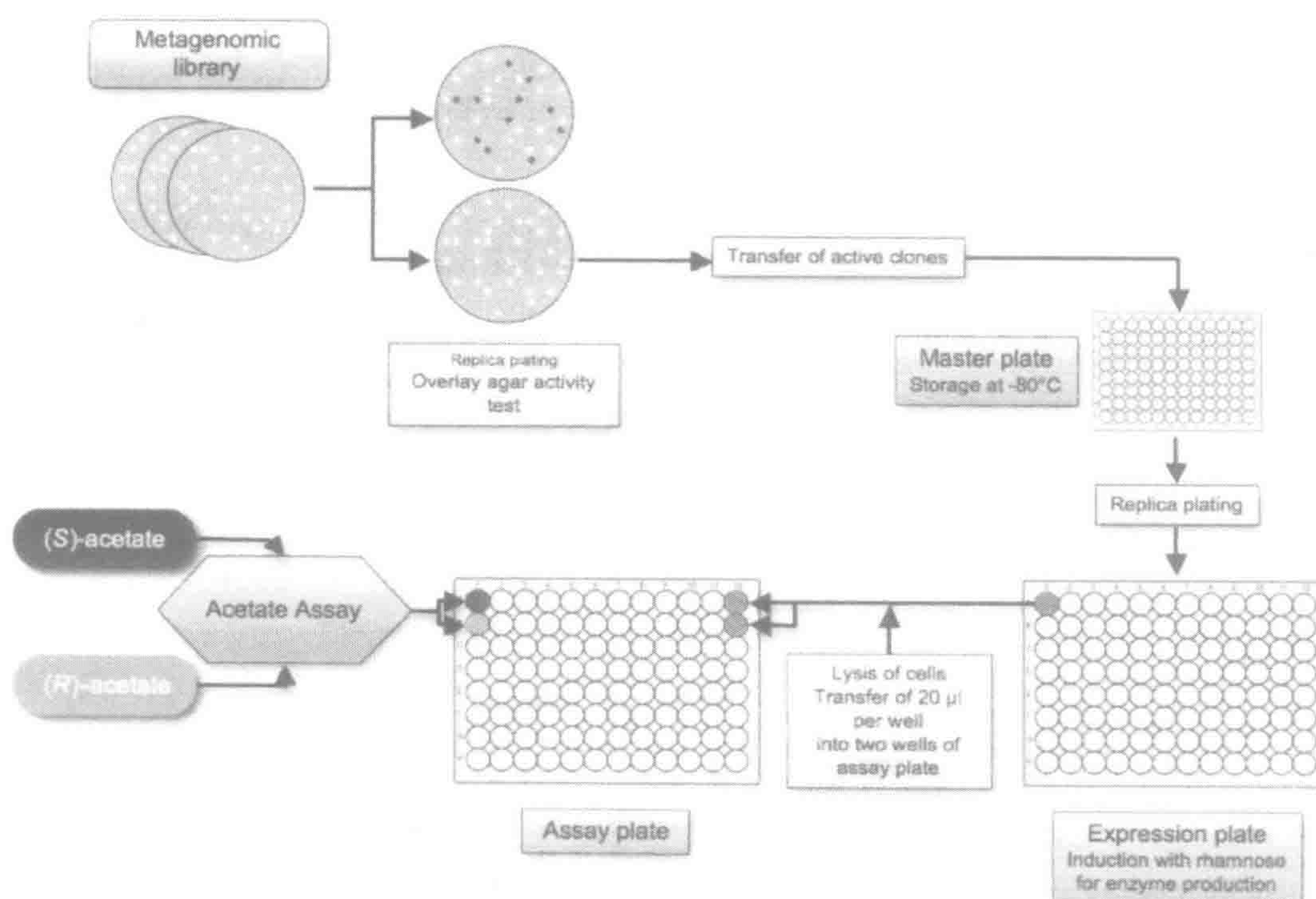


Fig. 2. Enzyme production in microtiter plates and principle of the screening of metagenomic libraries for altered enantioselectivity by adding (*R*)- and (*S*)-substrates to separate wells of a microtiter plate containing the same enzyme variant. If only activity is measured, the enzyme sample must not be split into two wells.

3.3. Cell Lysis in Microtiter Plates

1. After cultivation for approximately 5 h at 37°C and 220 rpm, centrifuge at 2,000 × *g* for 15 min, discard the supernatant, and add 200 μL lysis buffer.
2. Incubate the plates for 30 min at 4°C, freeze the plates at –80°C for 1 h, and thaw them at 37°C for approximately 20 min.
3. Centrifuge again at 2,000 × *g* for 15 min and transfer enzyme solution to new microtiter plates (see Note 3).

3.4. Screening for Activity or Enantioselectivity

1. To a mixture of the test-kit components (150 μL), add 20 μL enzyme solution (see Note 4) from the production plate. Either an activity test (go on with step 1 option A) or selectivity test (go on with step 1 option B) can be performed.

A: Activity Test

1. Start the reaction by adding 20 μL of substrate solution [i.e., racemic acetic acid esters, substrate concentration 5–50 mM dissolved in sodium phosphate buffer (10 mM, pH 7.4)].
2. Monitor the increase of NADH at 340 nm over 10 min (see Note 5). Use mixtures of the test kit with cell lysates of *Escherichia coli* harboring the empty expression vector without enzyme-encoding gene as negative control (see Note 6). A positive control is included in the test kit.

B: Selectivity Test

1. Transfer 20 μL enzyme solution from one well into two wells of a new microtiter plate (Fig. 2).
2. Add 20 μL of optically pure (*R*)- or (*S*)-acetates alternately into the rows of the plate.
3. Measure the increase of absorption at 340 nm for 10 min (see Note 5) and calculate the initial reaction rates for each enantiomer separately. The quotient of these rates is the apparent enantioselectivity E_{app} (Fig. 3).

4. Notes

1. For higher accuracy, all pipetting steps should be done by a pipetting robot.
2. Dilution with fresh medium is very important to achieve a comparable cell density in each well of the microtiter plate.
3. These plates can be stored at –20°C, freeze-dried, or directly used for the determination of enantioselectivity/activity. Freeze-dry the enzyme solution if you are expecting only very low

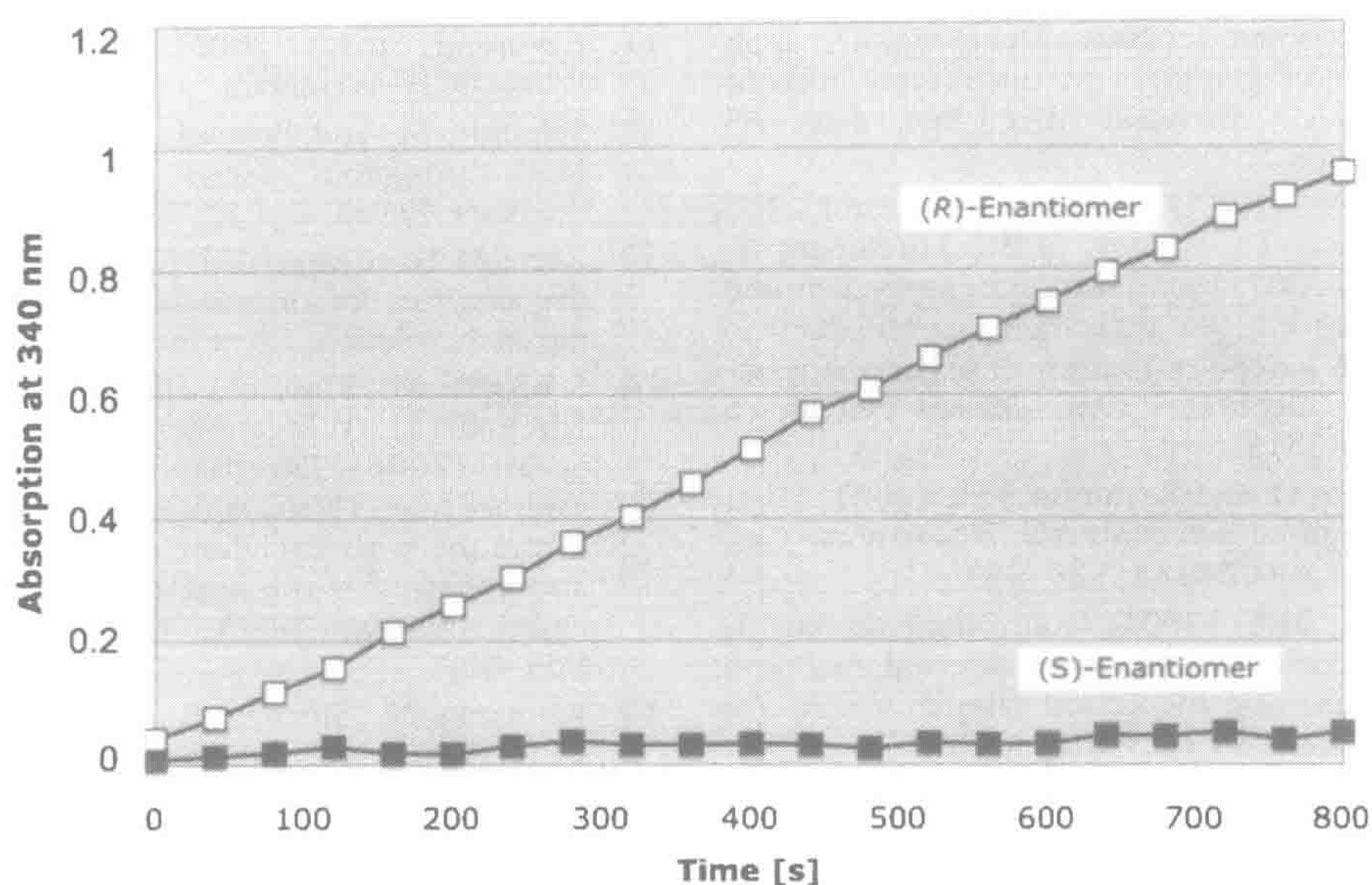


Fig. 3. Initial rates determined by using optically pure (*R*)- or (*S*)-1-phenyl ethyl acetate (5 mg/μL). Reactions were performed by using lyophilized crude cell extract of PFE.

activity in the diluted cell extract. Make sure that the enzyme tolerates this procedure.

4. If you are using freeze-dried enzyme, dissolve it first with 200 μL sodium phosphate buffer per well of the microtiter plate and then transfer 20 μL to each well of the assay plate.
5. Make sure that the absorption increase is in the linear range. In case of nonlinear behavior, the enzyme might be too active and dilution of enzyme solution in a new microtiter plate should solve this problem.
6. Make sure by running appropriate control reactions that no acetic acid is present prior to substrate (acetate) addition.
7. If no activity is measured, an increase of the enzyme amount or prolonged reaction time may exclude false-negative results.

References

1. Bornscheuer, U.T. and Kazlauskas, R.J. (2006) Hydrolases in Organic Synthesis – Regio- and Stereoselective Biotransformations, 2nd Edn. Wiley-VCH, Weinheim.
2. Cherry, J.R. and Fidantsef, A.L. (2003) Directed evolution of industrial enzymes: an update. *Curr Opin Biotechnol* 14, 438–443.
3. Turner, N.J. (2003) Directed evolution of enzymes for applied biocatalysis. *Trends Biotechnol* 21, 474–478.
4. Schmidt, M., Baumann, M., Henke, E., Konarzycka-Bessler, M., and Bornscheuer, U.T. (2004) Directed evolution of lipases and esterases. *Methods Enzymol* 388, 199–207.
5. Lorenz, P., Liebeton, K., Niehaus, F., and Eck, J. (2002) Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. *Curr Opin Biotechnol* 13, 572–577.

6. Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 68, 669–685.
7. Kourist, R., Krishna, S.H., Patel, J.S., Bartnek, F., Weiner, D.W., Hitchman, T., et al. (2007) Identification of a metagenome-derived esterase with high enantioselectivity in the kinetic resolution of arylaliphatic tertiary alcohols. *Org Biomol Chem* 5, 3310–3313.
8. Wahler, D. and Reymond, J.-L. (2001) High-throughput screening for biocatalysts. *Curr Opin Chem Biol* 12, 535–544.
9. Reetz, M.T. (2004) Controlling the enantioselectivity of enzymes by directed evolution: practical and theoretical ramnifications. *Proc Natl Acad Sci USA* 101, 5716–5722.
10. Goddard, J.P. and Reymond, J.-L. (2004) Recent advances in enzyme assays. *Trends Biotechnol* 22, 363–370.
11. Reymond, J.-L. (2005) Enzyme Assays. Wiley-VCH, Weinheim.
12. Schmidt, M. and Bornscheuer, U.T. (2005) High-throughput assays for lipases and esterases. *Biomol Eng* 22, 51–56.
13. Reetz, M.T. and Jaeger, K.E. (2000) Enantioselective enzymes for organic synthesis created by directed evolution. *Chem Eur J* 6, 407–412.
14. Schmidt, M., Hasenpusch, D., Kähler, M., Kirchner, U., Wiggenhorn, K., Langel, W., et al. (2006) Directed evolution of an esterase from *Pseudomonas fluorescens* yields a mutant with excellent enantioselectivity and activity for the kinetic resolution of a chiral building block. *ChemBioChem* 7, 805–809.
15. Baumann, M., Stürmer, R., and Bornscheuer, U.T. (2001) A high-throughput-screening method for the identification of active and enantioselective hydrolases. *Angew Chem Int Ed Engl* 40, 4201–4204.

Chapter 12

Screening for Cellulase-Encoding Clones in Metagenomic Libraries

Nele Ilmberger and Wolfgang R. Streit

Abstract

Modern biotechnology has the steady need to continuously identify novel enzymes for use in biotechnological applications. In industrial applications, however, enzymes often have to function under extreme and nonnatural conditions (i.e., in the presence of solvents, high temperature and/or at extreme pH values). Cellulases have many industrial applications from the generation of bioethanol, a realistic long-term energy source, to the finishing of textiles. These industrial processes require cellulolytic activity under a range of pH, temperature, and ionic conditions, and they are usually carried out by mixtures of cellulases. Investigation of the broad diversity of cellulolytic enzymes involved in the natural degradation of cellulose is necessary for optimization of these processes.

Key words: Cellulase, Ionic liquid, Metagenome, Bioethanol, Renewable energy, Biotechnology

1. Introduction

Metagenomics has recently become a very powerful tool to search for novel enzymes that are useful for biotechnological applications. A number of reviews have summarized the technology (1–3). Since its first publication and the description of the basic technology (4), a remarkable number of reports were published, providing new enzymes with a high potential for industrial applications (5–8). Because cellulose is a valuable biopolymer for the production of biofuels (i.e., ethanol) and other bio-based products, a significant number of publications have been reported on the isolation of metagenome-derived cellulases. Functional screening of a soil metagenomic library for cellulases revealed a total of eight cellulolytic clones, one of which was purified and characterized (9). Metagenomic screening of extreme environments, soda lakes in Africa and Egypt, detected more than a dozen cellulases, some of

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668,
DOI 10.1007/978-1-60761-823-2_12,

which displayed habitat-related halotolerant characteristics (10, 11). One of the earliest articles presenting metagenome-derived biocatalysts reported the detection of cellulases from a thermophilic, anaerobic digester fueled by lignocellulose (12), and a very recent study has detected seven cellulases with novel features (13). While most metagenomic surveys for novel cellulases concentrate on extreme environments, there is sufficient evidence that nonextreme, and therefore highly genetically diverse, environments also contain a range of cellulases that are highly stable and suitable for industrial applications (9, 14). Further examples of successful isolation of metagenome-derived cellulases have been described (15, 16). It is noteworthy that the sequencing-based approaches in diverse metagenomes have led to the identification of numerous putative cellulose clones (17). Of course, the functionality of these enzymes has to be verified.

Cellulose is, next to chitin, probably the most abundant renewable energy source; plants usually contain 35–50% (dry weight) cellulose. It can be used as a valuable source for bioethanol and other products. Therefore cellulose, consisting of β-1,4-linked glucose subunits (Fig. 1), must be hydrolyzed into fermentable sugar. Breakdown of cellulose can be performed by chemical treatment or enzymatic hydrolysis. Chemical breakdown has the disadvantage of cost-intensive pollutants. For large-scale enzymatic hydrolysis, a problem arises: cellulose is water-insoluble, while cellulases need an aquatic environment for their functionality. One solution might be the use of ionic liquids as solvent. These are salts liquid at room temperature, have no detectable vapor pressure, and are recyclable. Additionally, some ionic liquids have been described to dissolve cellulose (18–20).

Cellulases are distinguished from other glycoside hydrolases by their ability to hydrolyze ß-1,4-glucosidic bonds between glucosyl residues. The enzymatic breakage of the ß-1,4-glucosidic bonds in cellulose proceeds through an acid hydrolysis mechanism using a proton donor and nucleophile or base. The hydrolysis can either result in the inversion or retention (double replacement mechanism) of the anomeric configuration of carbon-1 at the reducing end (21–23).

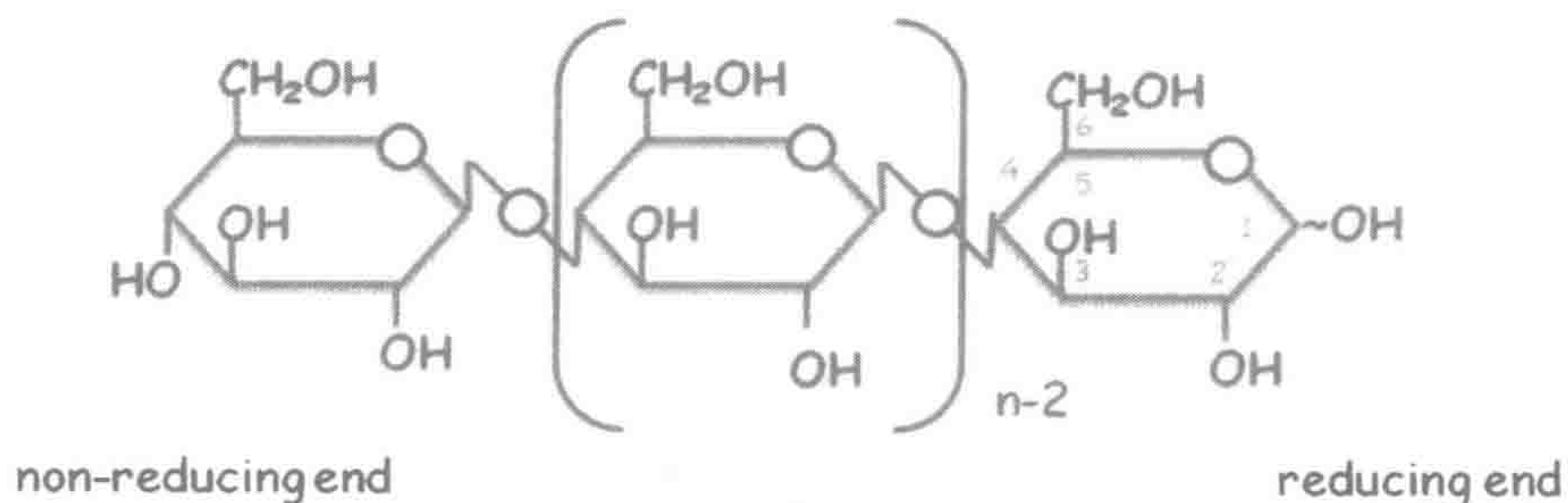


Fig. 1. Cellulose structure D-glucose; linked to large polymers via the ß-1,4 glycosidic linkage.

Three major types of enzymatic activity are necessary for complete degradation of cellulose: endoglucanases (1,4-ß-D-glucan-4-glucanohydrolases; EC 3.2.1.4), exoglucanases including cellodextrinases (1,4-ß-D-glucan glucanohydrolases; EC 3.2.1.74) and cellobiohydrolases (1,4-ß-D-glucan cellobiohydrolases; EC 3.2.1.91) and ß-glucosidases (ß-glucoside glucohydrolases; EC 3.2.1.21) (24, 25). The most recent nomenclature describes more than 100 families of glycosyl hydrolases, which are organized into 14 clans as listed at the CAZy server (http://afmb.cnrs-mrs.fr/CAZY/).

Cellulases have many industrial applications; next to the generation of bioethanol, a realistic long-term energy source, e.g. the finishing of textiles (26, 27). These industrial processes require cellulolytic activity under a variety of pH, temperature, and ionic conditions, and they are usually carried out by mixtures of cellulases. Investigation of the broad diversity of cellulolytic enzymes involved in the natural degradation of cellulose is necessary for the optimization of these processes.

While there remains much interest in the isolation of cellulases from fungal sources, there has been a recent increase in the isolation of diverse novel cellulases from prokaryotic organisms (23). The two different structural types of cellulolytic systems found in bacteria are noncomplexed and complexed systems. Some anaerobes are known to produce an extracellular multienzyme complex called cellulosome, which is linked to the cell surface (28). A cellulosome comprises different hydrolases organized on a noncatalytic scaffolding protein that mediates the attachment to cellulose (28). In contrast, cellulases from the majority of aerobes are not organized as complexes but bind directly to cellulose (29). These noncomplexed cellulases can have a modular structure with non-catalytic carbohydrate binding domains (CBD) and other domains like Ig-like domains connected to the catalytic domain(s) by flexible linkers. CBDs play a role in binding the cellulase to insoluble cellulose (30, 31). In addition to enzymes with clearly designated carbohydrate-binding domains, a significant number of cellulases have been identified that have no stated CBD and are thus referred to as nonmodular cellulases (27). Cellulases lacking a CBD show reduced activities against insoluble cellulose while retaining the capacity to depolymerize soluble cellulosic substrates (30, 32, 33).

The majority of the so far investigated prokaryotic cellulases have been isolated from cultured microorganisms. Cellulases from specific organisms tend to be active in the pH and temperature conditions corresponding to their environment such as the ß-1,4-endoglucanase from the gut bacterium *Cellulomonas pachnodae*, which has a pH range between pH 4.8 and 6.0 (34), and the endoglucanase from an alkalophilic *Bacillus* species, which has a pH range from pH to 12.0 (35). Industrial purposes require enzymes that are stable and active under specific conditions of pH,

temperature, and ionic strength. Many of the cellulases with the industrially relevant characteristics are obtained from extremophile microorganisms (26, 36). Cultivation of microbes from these or other specific environments is particularly problematic, what results in a large proportion of uncultured bacteria, especially in these habitats. Metagenomics is a cultivation independent analysis of the microbial DNA of a specific habitat and involves direct isolation of DNA from the environment followed by cloning and expression of the metagenome in a heterologous host (37). This technique has been used to detect a wide range of biocatalysts from uncultured microorganisms (1, 3). Here, we offer some easy-to-follow protocols for screening microbial cellulases in metagenomes.

2. Materials

2.1. Mineral Salt Medium (MSM)

1. Solution 1 (1 L, 10×): 70 g $Na_2HPO_4 \times 2H_2O$, 20 g KH_2PO_4.
2. Solution 2 (1 L, 10×): 10 g $(NH_4)_2SO_4$, 2 g $MgCl_2 \times 6H_2O$, 1 g $Ca(NO_3)_2 \times 4H_2O$.
3. Trace elements (2,000×, 1 L): 5 g EDTA, 3 g Fe(III) $SO_4 \times 7H_2O$, 30 mg $MnCl_2 \times 4H_2O$, 50 mg $CoCl_2 \times 6H_2O$, 20 mg $NiCl_2 \times 2H_2O$, 10 mg $CuCl_2 \times 2H_2O$, 30 mg $Na_2MoO_4 \times 2H_2O$, 50 mg $ZnSO_4 \times 7H_2O$, 20 mg H_3BO_4, pH 4.0.
4. Vitamins (1,000×, 100 mL): 1 mg biotin, 10 mg nicotinic acid, 10 mg thiamin–HCl (vitamin B1), 1 mg p-aminobenzoic acid, 10 mg Ca–D(+) pantothenic acid, 10 mg vitamin B6 hydrochloride, 10 mg vit. B12, 10 mg riboflavin, 1 mg folic acid.

2.2. Congo Red Plate Assay

1. LB-Agar + carboxymethylcellulose (CMC) (1 L): 15 g agar, 10 g tryptone, 5 g yeast extract, 5 g NaCl, 2 g CMC.
2. Congo red solution: 0.2% congo red.

2.3. DNSA-Assay

1. LB + CMC (1 L): 10 g tryptone, 5 g yeast extract, 5 g NaCl, 2 g CMC.
2. DNSA-reagent (1 L): 10 g 3,5-dinitrosalicylic acid, 2 mL phenol, 0.5 g Na_2SO_3, 200 g K-Na-tartrat, 10 g NaOH. Store at 4°C (protected from light).
3. McIllvaine-buffer: 0.2 M Na_2HPO_4 (A), 0.1 M citric acid (B). pH 6.5 is adjusted by the addition of (B) to (A) at 65°C.

2.4. Primers for 16S/18S PCR Proposed in this Chapter are

1. Archaea 20F: TTC CGG TTG ATC CGC CRG
   927R: TCC GGC GTT GAM TCA ATT.
2. Bacteria 616V: AGA GTT TGA TYM TGG CTC AG (38)
   1492R: CGG YTA CCT TGT TAC GAC (39).

3. Eucarya (18S) E4: AGG AAT TGA CGG AGG GCA C
   E1688: GGA CAT CTA AGG GAT CAC A.

2.5. Analysis of Cellulase Reaction Products by Thin-Layer Chromatography

1. Used substrates might be: cellooligosaccharides (1%, Sigma, Heidelberg, Germany), lichenan (1%, from *Cetraria islandica*, Sigma, Heidelberg, Germany), and CMC (1%, Sigma, Heidelberg, Germany).
2. Cellulase extract in 50 mM K_2HPO_4.
3. Silica 60 TLC plate (Merck KGaA, Darmstadt, Germany).
4. 1-Propanol, nitromethane, H_2O (5:3:2, vol/vol/vol).
5. Ethanol/concentrated sulphuric acid (9:1, vol/vol). Prepare fresh.
6. Ethylacetate, acetic acid, H_2O (2:1:1, vol/vol/vol).
7. Phosphoric acid.
8. Stock solution: 1 g diphenylamine, 1 mL aniline, 100 mL acetone.
9. 1-Propanol, ethylacetate, H_2O (6:1:3, vol/vol/vol).

2.6. Analysis of Cellulose Breakdown Products by HPLC

1. SepPack cartridge 18 (Waters, Milford, MA).
2. HPX-42A carbohydrate column (300×7.8 mm; BioRad, Munich, Germany).
3. Differential refractometer.

3. Methods

3.1. Enrichment of Highly Cellulolytic Microbial Communities

3.1.1. Enrichment Cultures (see Note 1)

In our experience, the number of clones that encode cellulases in environmental libraries is rather low. Therefore it is sometimes useful to slightly enrich on a suitable substrate to increase the frequency of cellulolytic organisms and therefore enzymes. Therefore usually mineral salt media (see Subheading 2.1) are used. The cultures are run under the desired conditions of pH, temperature, oxygen supply, etc. For the enrichment of cellulolytic organisms, cellulosic substrates like CMC, crystalline cellulose like avicel, cellulosic filter paper or plant material like wood or silage can be used as sole carbon source.

Once microbial communities are established, they can be used for library construction. Library construction from the enriched consortia is similar to the library construction of nonenriched microbial communities. The protocols are given in Chapter 2. It is recommended to analyze the microbial community by 16S profiling to verify the diversity (see Subheading 3.1.2). Please note that due to the enrichment steps, the diversity is probably significantly reduced.

3.1.2. Verification of Biodiversity in Enriched Cultures

For the evaluation of enrichment cultures and other samples, a 16S rRNA analysis might be a reasonable step. PCR fragments will then be cloned into pDrive (Qiagen, Hilden, Germany) or any comparable AT-cloning vector. The resulting plasmids will be purified, e.g. with the plasmid isolation kit from Qiagen (Hilden, Germany) and sequenced. Sequences will be corrected according to the sequencing fluorescence curve. Sequences can be analyzed by using the ARB database (http://www.arb-home.de) or the ribosomal database (http://rdp.cme.msu.edu).

3.2. Identification of Cellulase-Positive Clones by Screening on Congo Red Plates (see Note 2)

Cellulase-positive clones are usually screened by using a colorimetric assay on plates containing a cellulosic substrate. The interaction of the direct dye congo red with intact β-D-glucans provides the basis for a rapid and sensitive screening test for cellulolytic bacteria possessing β-D-glucan-hydrolase activities (40). The *E. coli* clones are stamped or streaked on LB agar with 0.2% CMC as substrate and incubated overnight at 37°C, followed by an incubation of 2–7 days at RT. Colonies are washed off with ddH_2O to permit the homogeneous penetration of the staining dye into the medium. Agar plates are stained with congo red solution for 30 min. The solution is poured off and the agar plates are destained up to three times for 30 min with 1 M NaCl. Cellulase-expressing clones are detected by the formation of a yellow halo against a red background (see Fig. 2).

3.3. Retransformation of Putative Positive Clones

To ensure that the observed catalytic activity of clones is not due to contaminations, the isolation and retransformation of the vector and a subsequent activity assay are recommended. Only then clones should be stored and used for further work.

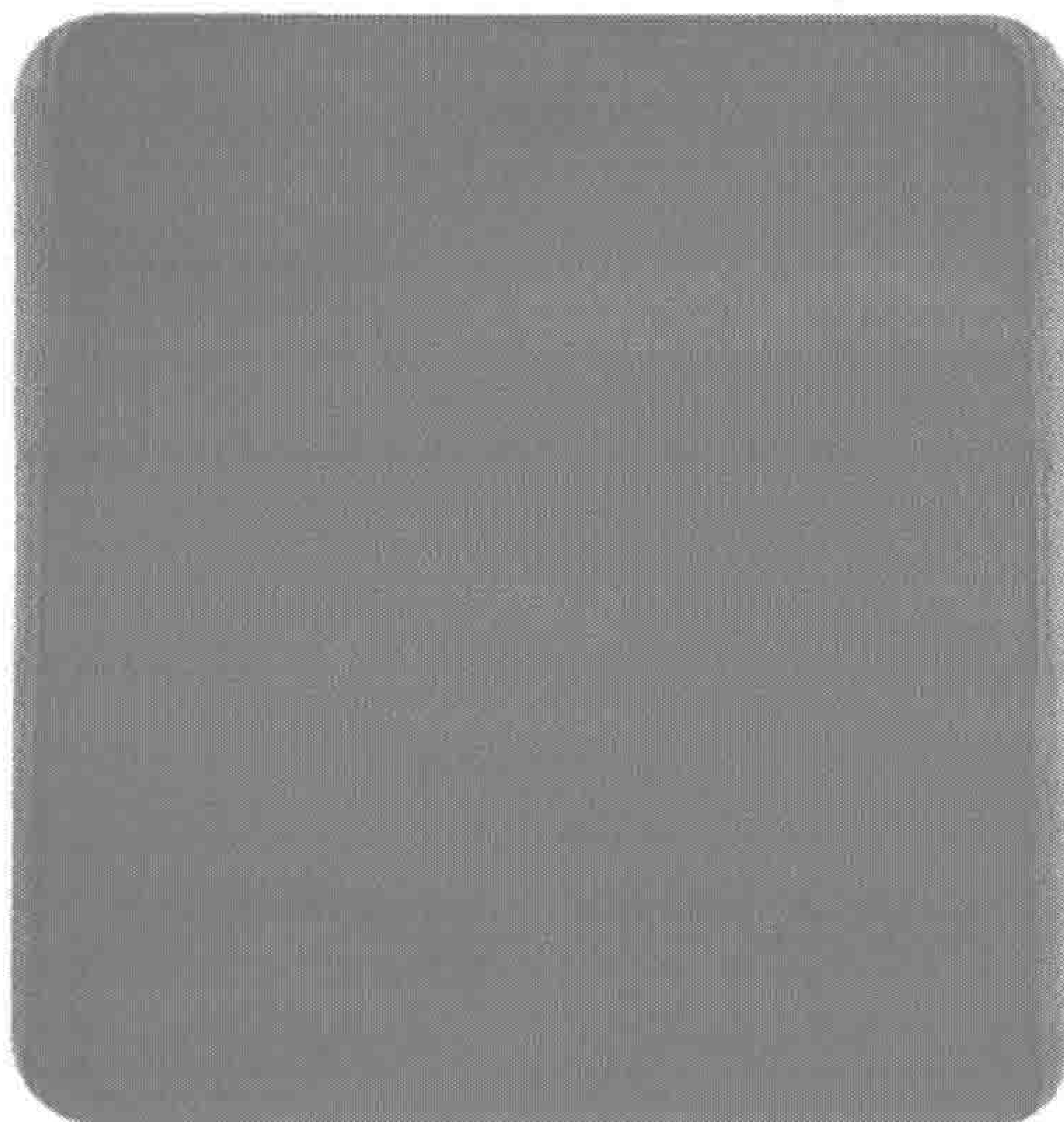


Fig. 2. Activity staining of metagenome-derived cosmid clones using *congo red staining*.

3.4. Preparation of Crude Cell Extracts of Clones with Cellulolytic Activity

For the preparation of crude cell extracts of cellulase-positive clones, 200 mL LB cultures with 0.2% (w/v) CMC containing an appropriate antibiotic are grown at 37°C to an OD of 1.0–1.5. Cells are harvested and resuspended in 50 mM Tris–HCl pH 8.0 prior to cell disruption through sonication (Sonicator UP 200S, Hielscher, Germany) at 50% amplitude and cycle 0.5 for 5 min. After centrifuging at 16,000 × g at 4°C for 30 min, the crude cell extract can be stored at 4°C for several days.

3.5. Enzyme Assays for Cellulase Activities

3.5.1. DNSA Assay (see Note 3)

Cellulase activity is routinely assayed by measuring the amount of reducing sugar released from CMC using 3,5-dinitrosalicylic acid reagent (see Subheading 2.3). The standard assay mixture contains 2 μg of the enzyme or crude cell extract and 1% CMC in a final volume of 0.5 mL with 150 μL McIllvaine buffer (see Subheading 2.3). This mixture is incubated at an appropriate temperature (usually 37°C) for 15 min. By the hydrolysis of cellulose, glucose oligomers and monomers are produced. By this process the number of reducing ends increases. These reducing groups react with 3,5-dinitrosalicylic acid forming brown 3-amino-5-nitrosalicylic acid at 100°C. The amount of 3-amino-5-nitrosalicylic acid formed is equimolar to the number of reducing ends. Therefore the amount of reducing sugars can be quantified at 546 nm (Fig. 3).

Units of enzyme activity (U) are expressed as micromoles of reducing sugar released per minute per milligram protein. Enzyme activities are formulated by regressing absorbance on concentration following the Beer's law. That is the relationship between known concentrations, and absorbance is linear except at very low or high concentration of the product, in this case reducing sugar. One unit is equal to 1 μmol of reduced sugar per minute. The enzymatic activity volume was calculated according to the following formula:

$$\mathrm{U/mL} = (\Delta E/\mathrm{min} \times V)/(\varepsilon \times d \times v)$$

$\Delta E/\mathrm{min}$ = extinction
V = volume of the test reaction mix.
D = thickness of the cuvette [cm].
ε = ascendant of straight calibration line.
v = sample volume.

The specific enzymatic activity [U/mg protein] is defined as the amount of enzyme that liberates 1 μmol of substrate per minute and is calculated as follows.

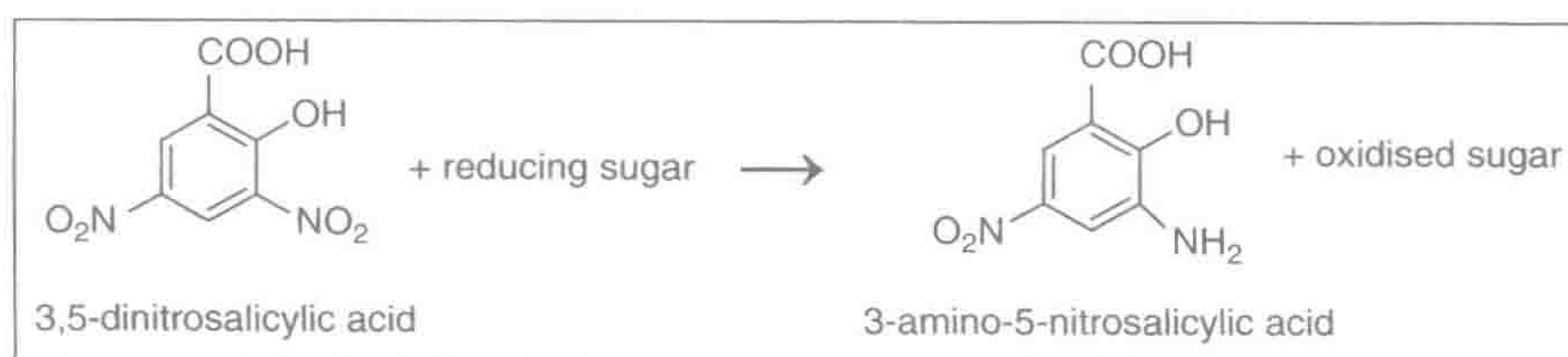


Fig. 3. DNSA assay reaction for the measurement of cellulolytic activity based on the release of reducing sugar ends.

Specific activity [U/mg protein] = enzymatic activity volume [U/mL]/protein concentration [mg/mL]

The reaction mixes are prepared by combining first buffer and enzyme and then adding the substrate. The mixture is incubated for 15 min at 37°C. After this incubation, DNSA reagent is added and the samples are boiled at 100°C for 15 min. After cooling down on ice, the samples are centrifuged at 16,000 × *g* for 2 min to precipitate falling proteins. The samples are transferred to cuvettes and absorbance was measured at 546 nm.

The pH range of the enzyme is usually determined by measuring standard assay activity between pH 4 and 10.5 using 50 mM of appropriate buffers. Acetate buffer is used for pH 4–6.0, citrate/phosphate buffer (McIllvaine buffer) is used for pH 6–7.5, Tris–HCl is used for pH 7.5–9.0 and *N*-cyclohexyl-3-aminopropanesulfonic acid (CAPS) is used for pH 9.7–10.5. For the analysis of the temperature range of the enzyme, activity of the standard assay mixture is assayed at temperatures between 20 and 95°C. To analyze substrate specificity, CMC can be replaced in the standard assay mixture with lichenan, barley β-glucan, laminarin, oat-spelt xylan or avicel.

Reaction mix:

Sample	100 µL
CMC in ddH_2O (2%)	250 µL
McIllvaine-buffer, pH 6.5	*150* µL
DNSA reagent	750 µL

Inhibition or enhancement of cellulase activity can be determined for a range of different metal chloride salts, solvents, detergents, and EDTA using in general 1 mM concentrations. The influence of ionic liquids (IL) can be evaluated in the standard assay mixture system when McIllvaine (see Subheading 2.3) buffer is replaced by an IL. The assay mixture, therefore, comprises an IL content of 30%. (ILs that can be used for cellulase activity assays are depicted in Fig. 4) This value can be up- and down-regulated. For ILs as well as other additives long-term stability assays might be of interest. Therefore, the enzyme is incubated in buffer with the desired additives at the favored conditions for different time periods. Then the substrate is added and the assay proceeded as described above.

3.5.2. Analysis of Cellulase Reaction Products by Thin-Layer Chromatography (TLC) (see Note 4)

To determine whether a cellulase has an endo or exo mode of action, TLC analyses are an adequate tool. These analyses can also give a good overview on the substrate range hydrolyzed by the enzyme.

As substrates, different carbohydrates can be used, e.g., cellooligosaccharides, lichenan, and CMC. These substrates are coincubated with cellulase extract in 50 mM K_2HPO_4 at adequate pH and temperature conditions. To determine which reaction products

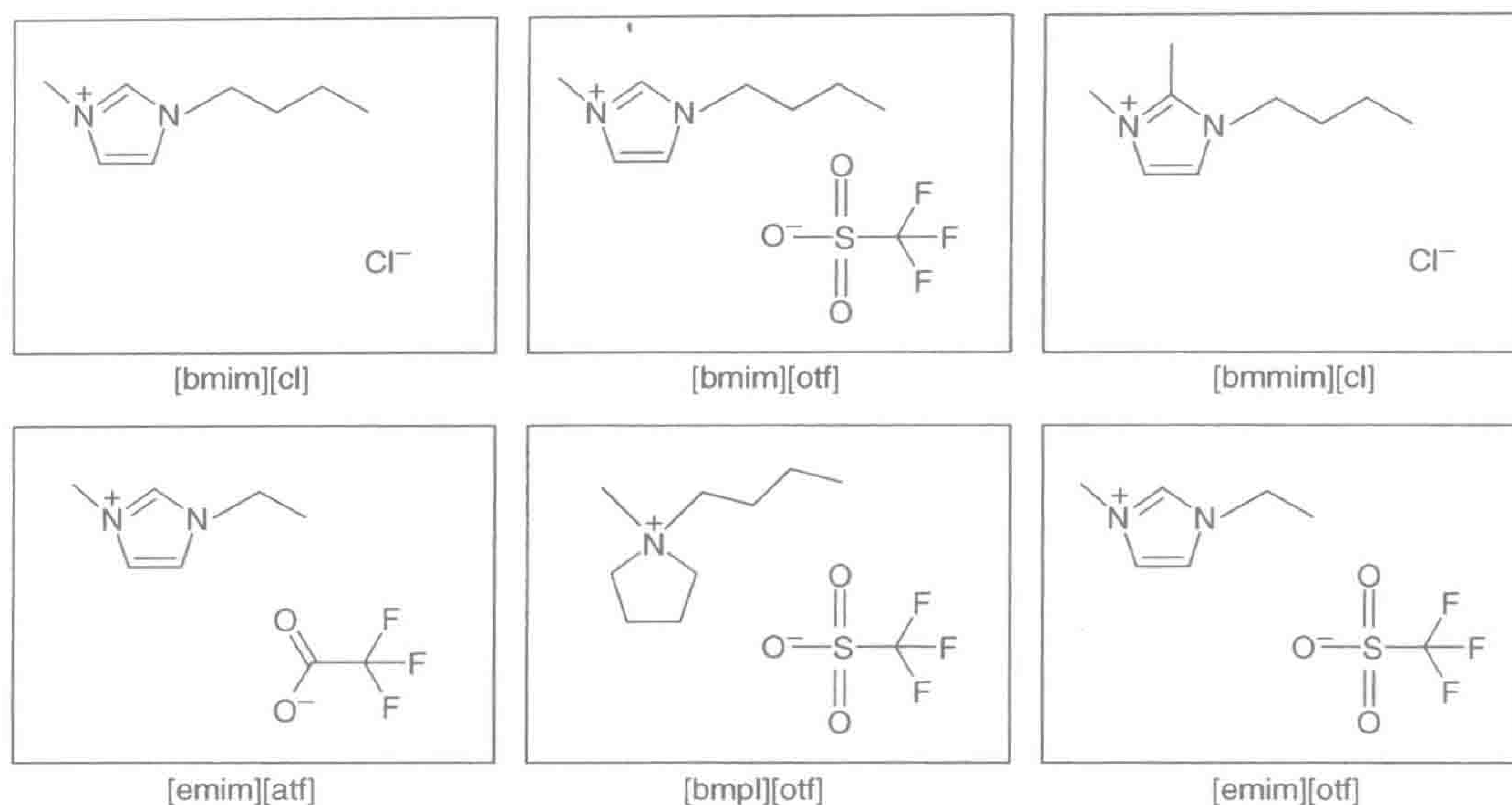


Fig. 4. Ionic liquids that are suitable for cellulase activity assays.

occur first, aliquots from different incubation times can be spotted on a silica 60 TLC plate. The cellooligosaccharide reaction products are developed and separated in 1-propanol, nitromethane, H_2O (5:3:2, vol/vol/vol) for 2 h. After separation, sugars are visualized by spraying the plates with a freshly prepared mixture of ethanol/concentrated sulphuric acid (9:1, vol/vol). The lichenan reaction products are developed in ethylacetate, acetic acid, H_2O (2:1:1, vol/vol/vol) for 3 h. After separation, sugars are visualized by spraying the plates with a freshly prepared mixture of 1 mL phosphoric acid and 10 mL stock solution (1 g diphenylamine, 1 mL aniline, 100 mL acetone). The CMC reaction products are separated and developed in 1-propanol, ethylacetate, H_2O (6:1:3, vol/vol/vol) for 2 × 3 h, and the sugars visualized using the same mixture as for visualization of lichenan products.

3.5.3. Analysis of Cellulose Breakdown Products by HPLC (see Note 4)

For the investigation of the reaction products of hydrolysis of carbohydrates, HPCL analysis is an appropriate method (Fig. 5). First, enzyme preparation and substrate are coincubated for 2 h at optimal temperature and pH of McIllvaine buffer. As well as for TLC analysis and DNSA assays, different substrates and reaction conditions can be investigated. For stopping the reactions, assays were incubated at 100°C for 10 min. The assay mixtures were centrifuged and proteins of the supernatant were removed with a SepPack cartridge 18. There are a lot of different HPLC columns and elution buffers that can be used for the analysis of carbohydrate hydrolysis reaction products. One possibility is the analysis of the samples with a HPX-42A carbohydrate collumn. Elution was carried out with H_2O at 85°C; the flow rate was 0.6 mL/min. Detection was performed with a differential refractometer.

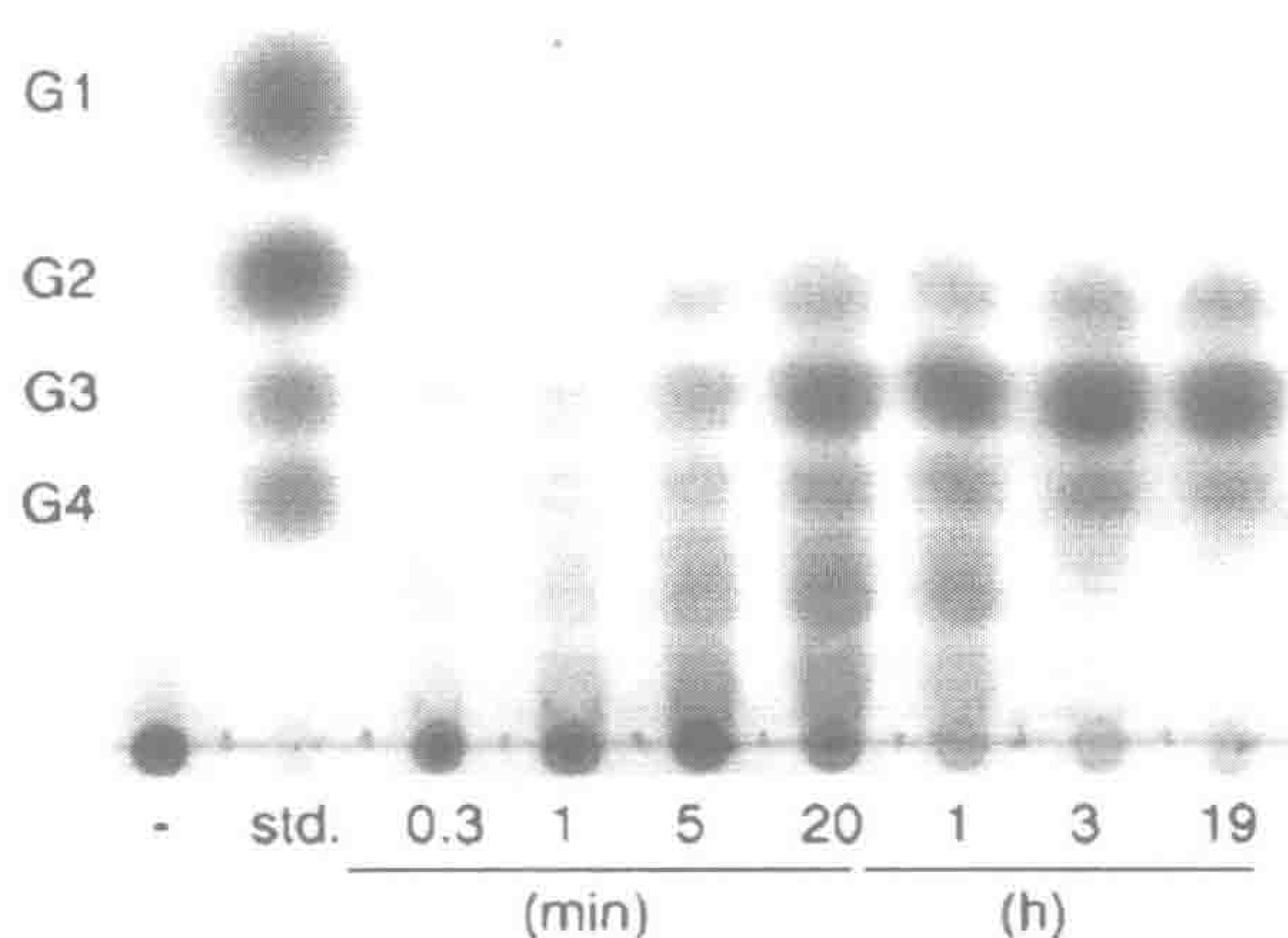


Fig. 5. TLC detection of end products from cellulose degradation: Lane (–) is the sample without the addition of enzyme, (std.) is the standard with glucose (G1), cellobiose (G2), cellotriose (G3), and cellotetraose (G4). The other lanes are different time points of the hydrolysis of lichenan with Cel5A (9).

4. Notes

Altogether, screening for and assaying cellulases is not that complicated, here we just announce some simple notes.

1. The most "critical" step in this procedure and for the discovery of a pool of enzymes that is adequate for the detection of one or more cellulases with interesting properties might be the choice of sample and the quality of the enrichment culture and metagenomic library, respectively. We suggest investigating habitats with a high potential of the occurrence of cellulolytic bacteria, like intestinal tracts of herbivores or rotting trees. If an enrichment step is desired or inevitable, it is reasonable to enrich over a rather short time period to keep diversity as broad as possible.
2. Screening for cellulase-active clones on congo red indicator plates is easy; only the time period for growth of bacteria and expression of cellulolytic activity might be variable. Washing off bacterial cells is critical when cellulolytic activity is rather low.
3. The same occurs for the DNSA assay, where gloves should be worn, and when samples are boiled, the lid should be stabilized to protect from spraying phenol (in DNSA solution, see Subheading 2.3). When ionic liquids are added to the assay mixtures, it is necessary to completely agitate IL and aquatic phase, otherwise results are falsified.
4. For TLC and HPLC analysis concentrations and time points are critical for exploitable results. Another issue concerning both HPLC and TLC is the amount of sample. Applying too much will lead to worthless results.

References

1. Streit, W.R. and Schmitz, R.A. (2004) Metagenomics – the key to the uncultured microbes. *Curr Opin Microbiol* 7, 492–498.
2. Daniel, R. (2004) The soil metagenome – a rich resource for the discovery of novel natural products. *Curr Opin Biotechnol* 15, 199–204.
3. Schmeisser, C., Steele, H., and Streit, W.R. (2007) Metagenomics, biotechnology with nonculturable microbes. *Appl Microbiol Biotechnol* 75, 955–962.
4. Schmidt, T.M., DeLong, E.F., and Pace, N.R. (1991) Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J Bacteriol* 173, 4371–4378.
5. Ferrer, M., Golyshina, O.V., Chernikova, T.N., Khachane, A.N., Reyes-Duarte, D., Santos, V.A., et al. (2005) Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environ Microbiol* 7, 1996–2010.
6. Ferrer, M., Golyshina, O.V., Plou, F.J., Timmis, K.N., and Golyshin, P.N. (2005) A novel alpha-glucosidase from the acidophilic archaeon *Ferroplasma acidiphilum* strain Y with high transglycosylation activity and an unusual catalytic nucleophile. *Biochem J* 391, 269–276.
7. Beloqui, A., Pita, M., Polaina, J., Martinez-Arias, A., Golyshina, O.V., Zumarraga, M., et al. (2006) Novel polyphenol oxidase mined from a metagenome expression library of bovine rumen: biochemical properties, structural analysis, and phylogenetic relationships. *J Biol Chem* 281, 22933–22942.
8. Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K.-E., and Streit, W.R. (2003) Prospecting for novel biocatalysts in a soil metagenome. *Appl Environ Microbiol* 69, 6235–6242.
9. Voget, S., Steele, H.L., and Streit, W.R. (2006) Characterization of a metagenome-derived halotolerant cellulase. *J Biotechnol* 126, 26–36.
10. Grant, S., Sorokin, D.Y., Grant, W.D., Jones, B.E., and Heaphy, S. (2004) A phylogenetic analysis of Wadi el Natrun soda lake cellulase enrichment cultures and identification of cellulase genes from these cultures. *Extremophiles* 8, 421–429.
11. Rees, H.C., Grant, S., Jones, B., Grant, W.D., and Heaphy, S. (2003) Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. *Extremophiles* 7, 415–421.
12. Healy, F.G., Ray, R.M., Aldrich, H.C., Wilkie, A.C., Ingram, L.O., and Shanmugam, K.T. (1995) Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Appl Microbiol Biotechnol* 43, 667–674.
13. Feng, Y., Duan, C.J., Pang, H., Mo, X.C., Wu, C.F., Yu, Y., et al. (2007) Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. *Appl Microbiol Biotechnol* 75, 319–328.
14. Pottkämper, J., Barthen, P., Ilmberger, N., Schwaneberg, U., Schenk, A., Schulte, M., et al. (2009) Applying metagenomics for the identification of bacterial cellulases that are stable in ionic liquids. *Green Chem* 11, 957–965.
15. Guo, H., Feng, Y., Mo, X., Duan, C., Tang, J., and Feng, J. (2008) Cloning and expression of a beta-glucosidase gene umcel3G from metagenome of buffalo rumen and characterization of the translated product. *Sheng Wu Gong Cheng Xue Bao* 24, 232–238.
16. Pang, H., Zhang, P., Duan, C.J., Mo, X.C., Tang, J.L., and Feng, J.X. (2009) Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase. *Curr Microbiol* 58, 404–408.
17. Warnecke, F., Luginbuhl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450, 560–565.
18. Heinze, T., Schwikal, K., and Barthel, S. (2005) Ionic liquids as reaction medium in cellulose functionalization. *Macromol Biosci* 5, 520–525.
19. Swatloski, R.P., Spear, S.K., Holbrey, J.D., and Rogers, R.D. (2002) Dissolution of cellulose [correction of cellose] with ionic liquids. *J Am Chem Soc* 124, 4974–4975.
20. Wu, J., Zhang, J., Zhang, H., He, J., Ren, Q., and Guo, M. (2004) Homogeneous acetylation of cellulose in a new ionic liquid. *Biomacromolecules* 5, 266–268.
21. Beguin, P. and Aubert, J.P. (1994) The biological degradation of cellulose. *FEMS Microbiol Rev* 13, 25–58.
22. Birsan, C., Johnson, P., Joshi, M., MacLeod, A., McIntosh, L., Monem, V., et al. (1998) Mechanisms of cellulases and xylanases. *Biochem Soc Trans* 26, 156–160.
23. Hilden, L. and Johansson, G. (2004) Recent developments on cellulases and carbohydrate-binding modules with cellulose affinity. *Biotechnol Lett* 26, 1683–1693.

24. Bayer, E.A., Chanzy, H., Lamed, R., and Shoham, Y. (1998) Cellulose, cellulases and cellulosomes. *Curr Opin Struct Biol* 8, 548–557.
25. Kumar, R., Singh, S., and Singh, O.V. (2008) Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *J Ind Microbiol Biotechnol* 35, 377–391.
26. Ando, S., Ishida, H., Kosugi, Y., and Ishikawa, K. (2002) Hyperthermostable endoglucanase from *Pyrococcus horikoshii*. *Appl Environ Microbiol* 68, 430–433.
27. Lynd, L.R. and Zhang, Y. (2002) Quantitative determination of cellulase concentration as distinct from cell concentration in studies of microbial cellulose utilization: analytical framework and methodological approach. *Biotechnol Bioeng* 77, 467–475.
28. Schwarz, W.H. (2001) The cellulosome and cellulose degradation by anaerobic bacteria. *Appl Microbiol Biotechnol* 56, 634–649.
29. Zhang, Y.H. and Lynd, L.R. (2004) Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. *Biotechnol Bioeng* 88, 797–824.
30. Bolam, D.N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M.P., Rixon, J.E., et al. (1998) *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem J* 331 (Pt 3), 775–781.
31. Carvalho, A.L., Goyal, A., Prates, J.A., Bolam, D.N., Gilbert, H.J., Pires, V.M., et al. (2004) The family 11 carbohydrate-binding module of *Clostridium thermocellum* Lic26A-Cel5E accommodates beta-1,4- and beta-1,3-1,4-mixed linked glucans at a single binding site. *J Biol Chem* 279, 34785–34793.
32. Coutinho, J.B., Gilkes, N.R., Kilburn, D.G., Warren, R.A.J., and Miller, R.C., Jr. (1993) The nature of the cellulose-binding domain effects the activities of a bacterial endoglucanase on different forms of cellulose. *FEMS Microbiol Lett* 113, 211–217.
33. Fontes, C.M., Clarke, J.H., Hazlewood, G.P., Fernandes, T.H., Gilbert, H.J., and Ferreira, L.M. (1997) Possible roles for a non-modular, thermostable and proteinase-resistant cellulase from the mesophilic aerobic soil bacterium *Cellvibrio mixtus*. *Appl MicrobiolBiotechnol* 48, 473–479.
34. Cazemier, A.E., Verdoes, JC., Op den Camp, HJ., Hackstein, J.H., and van Ooyen, A.J. (1999) A beta-1,4-endoglucanase-encoding gene from *Cellulomonas pachnodae*. *Appl Microbiol Biotechnol* 52, 232–239.
35. Sanchez-Torres, J., Perez, P., and Santamaria, R.I. (1996) A cellulase gene from a new alkalophilic *Bacillus* sp. (strain N186-1). Its cloning, nucleotide sequence and expression in *Escherichia coli*. *Appl MicrobiolBiotechnol* 46, 149–155.
36. Solingen, P., Meijer, D., Kleij, W., Barnett, C., Bolle, R., Power, S., et al. (2001) Cloning and expression of an endocellulase gene from a novel streptomycete isolated from an East African soda lake. *Extremophiles* 5, 333.
37. Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M. (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5, R245–R249.
38. Brosius, J., Ullrich, A., Raker, M.A., Gray, A., Dull, T.J., Gutell, R.R., et al. (1981) Construction and fine mapping of recombinant plasmids containing the *rrnB* ribosomal RNA operon of *E. coli*. *Plasmid* 6, 112–118.
39. Kane, M.D., Poulsen, L.K., and Stahl, D.A. (1993) Monitoring the enrichment and isolation of sulfate-reducing bacteria by using oligonucleotide hybridization probes designed from environmentally derived 16S rRNA sequences. *Appl Environ Microbiol* 59, 682–686.
40. Teather, R.M. and Wood, P.J. (1982) Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Appl Environ Microbiol* 43, 777–780.

Chapter 13

Screening Metagenomic Libraries for Laccase Activities

Manuel Ferrer, Ana Beloqui, and Peter N. Golyshin

Abstract

Laccases are multi-copper oxidoreductases (benzenediol:oxygen oxidoreductases, EC 1.10.3.2) able to oxidise a wide variety of phenolic and non-phenolic compounds. They are useful enzymes for a variety of applications, including bioremediation and craft pulp bio-bleaching as the most significant ones. There is a considerable interest to find new laccases through the exploration of biological diversity. Laccases have been found in plants, insects, and bacteria but predominantly in fungi: these enzymes have been documented in about 60 fungal strains. Microbial diversity constitutes a largely unexplored treasure chest with new laccases with a good potential for basic science and biotechnology. At present, due to our inability to cultivate most microbes, the only means of accessing the resources of the microbial world is to harvest genetic resources ("metagenomes"), which can further on be subjected to extensive screening programs. In this chapter, we provide an overview of screening methods to identify laccase-encoding genes from environmental resources.

Key words: Laccase, Metagenome, Polyphenol oxidase, Screening, High throughput

1. Introduction

Laccases are multi-copper oxidoreductases (benzenediol:oxygen oxidoreductases, EC 1.10.3.2) that use molecular oxygen to oxidise a wide variety of phenolic and non-phenolic compounds, by a radical-catalysed reaction mechanism (1). Those compounds include industrial dyes, polycyclic aromatic hydrocarbons, pesticides, and alquenes. Laccases are also capable of performing polymerisation, depolymerisation, methylation, and demethylation reactions (2–5). Laccases are, therefore, useful enzymes for a variety of applications: decolourisation of different types of recalcitrant dyes (6), bioremediation of soils and water (7, 8), kraft pulp bio-bleaching (9), and few other biotechnological applications (7). They belong to a larger group of enzymes termed the

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668, DOI 10.1007/978-1-60761-823-2_13,

blue multi-copper oxidase family, which includes the plant ascorbate oxidase, the mammalian plasma protein ceruloplasmin, and bilirubin oxidase, among few others.

The term "laccase" stems from its original identification in the exudates of the Japanese lacquer tree *Rhus vernicifera* in 1883. Just over a century later, it was characterised as being a metal-containing oxidase (10). Laccases have also been found in other plants, animals, and bacteria but predominantly in fungi. Laccase activity has been demonstrated in more than 60 fungal strains belonging to *Basidiomycetes*, *Ascomycetes*, and *Deuteromycetes*, being documented in virtually every fungus examined for it (11). Its presence in plants appears to be far more limited than in fungi. All species of family *Anacardiaceae*, to which belongs the lacquer tree, contain laccase in the resin products and the secreted resin (12). Reports on the presence of laccase in other plants are so far limited to *Acer pseudoplatanus*, *Pinus taeda*, *Aesculus parviflora*, and *Populus eruamericana* (13); however, it is believed that they are present throughout the whole kingdom of *Plantae*. Polyphenol oxidases, perhaps laccase-like, have also been reported in insects (14), and there is strong evidence for the ubiquitous distribution of laccases in prokaryotes. The first bacterial laccase to be extensively studied was the one from *Axospirillum lipoferum*, and the crystal structure of a bacterial (*Bacillus subtilis*) laccase is now available (15). Sequence homology analysis suggests that laccases also occur in bacteria such as *Mycobacterium tuberculosis* (16). Laccase-encoding genes have been found in gram-negative and gram-positive bacteria, including species living in extreme habitats, e.g. in *Oceanobacillus iheyensis* and *Aquifex aeolicus* and in the archaeobacterium *Pyrobaculum aerophilum* (17, 18).

As described above, our current perception is extremely biased by, and mostly restricted to, laccases from organisms that are available in biochemically relevant quantities. However, much less than 1% of known microbes have thus far been cultured and are therefore missing on the whole biological picture, and it is conceivable that a high number of laccases remain undetected (19, 20). The harvesting and sequencing of environmental DNA – metagenomics – provide access to the genomic diversity of uncultured species, but not to their biochemistry. And, since the ability to annotate genes is based on sequence homologies, the new diversity revealed by metagenomics is restricted to revealing the micro-diversity within known protein families: new families representing new macro-diversity cannot be recognised. One means of sequence-independent exploration of wider diversity space is the creation and functional screening of metagenomic expression libraries, which identifies genes of enzymes independently of homology with any known gene. Following on from this, the use of metagenomics for laccase exploration has recently been assessed (20).

Using a bovine rumen microflora-derived library, a novel laccase was isolated and characterised. The laccase was unusual in three respects, namely, it lacked any sequence relatedness to the known laccases, it exhibited much higher activity and substrate affinities than any bacterial laccase thus far described, and it represented the first functionally characterised member of a new laccase family (apparently this protein turned out to be very ubiquitous in eubacteria). This clearly indicates that the screening of metagenomic libraries may lead in the future to the discovery of novel laccases, preferably of a bacterial origin. For that, screening methods should be defined in accordance with the known substrate specificity of known laccases.

1.1. Laccase Substrates

As mentioned above, laccases are widely distributed among prokaryotes and eukaryotes (13, 21, 22), and structural and comparative studies (23–25) have identified conserved regions in which histidine residues can bind four copper atoms located at two main sites that are involved in catalytic activity (5). The T1 copper site, which has the highest potential, is assumed to be the first electron acceptor. The other three copper ions form a cluster in two adjacent sites, designated T2/T3, one in the T2 site and two in the binuclear T3 site; electrons captured by the T1 site are transferred via T2/T3 to the product, leading to product oxidation and reduction of oxygen to water (5, 26).

Laccase catalyses the four electron reduction of molecular oxygen to water with one-electron oxidation of reducing substrate, without producing hydrogen peroxide. Although laccases preferably act on phenolic compounds, their substrate spectrum is extremely broad and strongly varies among laccases. In addition, there is an overlap in substrate range with another type of (copper-containing) oxidase – tyrosinase – notably a mono-phenol mono-oxygenase. Laccases can convert *o*- and *p*-diphenols, aminophenols, methoxy-substituted phenols, benzenethiols, polyphenols, polyamines, hydroxyindols, some aryl diamines, and a considerable range of other compounds, but do not oxidize tyrosine (whereas the tyrosinases do). Inorganic/organic metal compounds are also substrates of laccases. Mn^{2+} is oxidised to Mn^{3+} and $Fe(EDTA)^{2-}$ is also accepted by the enzyme. All known laccases catalyse the oxidation of ascorbic acid and phenol substrates with equally high efficiencies. Simple diphenols like hydroquinone and catechol are generally good substrates, but guaicol and 2,6-dimethoxyphenol are often better. *p*-Phenylene diamine is a frequently used substrate. Syringaldizine [*N,N'*-bis(3,5-dimethoxy-4-hydroxybenzylidene hydrazine); $\varepsilon_{525} = 65{,}000\ M^{-1}cm^{-1}$] is a good substrate, but it has to be used in the complete absence of hydrogen peroxide since syringaldizine is also oxidised by the manganese-dependent peroxidases produced by many lignolytic basidiomycetes (1, 11). Owing to its strong preference for many

laccases, it has successfully been employed for screening metagenomic libraries for laccase activity (20). Although differing in substrate specificity, laccases generally are more stable at alkaline pH than at acidic pH, probably due to the OH^- inhibition of auto-oxidation. Below, we recapitulate few different screening strategies, applicable in high-throughput format, used normally to detect laccase activity and which can be adapted to screen large expression libraries.

2. Materials

2.1. Laccase Substrates

1. 2,2′-Azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) (ABTS; ε_{420} = 38,000 $M^{-1}cm^{-1}$) (Sigma–Aldrich): working solutions are prepared from a 10 mM sodium acetate buffer (pH 4.5) stored at 4°C.
2. 2,6-Dimethoxyphenol (DMP; ε_{468} = 14,800 $M^{-1}cm^{-1}$), 3,4-dimetoxybenzyl alcohol (DMA; ε_{310} = 9,500 $M^{-1}cm^{-1}$), (2)4-metoxybenzyl alcohol (MBA; ε_{500} = 38,000 $M^{-1}cm^{-1}$) (Sigma–Aldrich): stock solution is prepared in methanol at a concentration of 200 μg/mL and kept at 4°C until use (see Note 1).
3. 1-Hydroxybenzotriazole (1-HBT; ε_{468} = 72.8 $M^{-1}cm^{-1}$) (Sigma–Aldrich): for preparation of 1-HBT stock dissolve 37 mg 1-HBT in 1.0 mL 0.4 M NaOH with stirring and add 1.0 mL citrate/Na_2H-phosphate buffer (0.06/0.08 M) pH 3.63 slowly to give pH 5.0–5.3 of the stock solution. Lower pH makes HBT to precipitate, while a higher pH affects the pH of the assay buffer. Store the stock 1-HBT solution in the dark at room temperature.
4. Syringaldazine (SGZ; ε_{530} = 64,000 $M^{-1}cm^{-1}$) (Sigma–Aldrich): SGZ, 0.56 mM stock solution: weigh 10.0 mg of SGZ in a weighing boat and transfer to the colorimetric flask. Add 96% ethanol to the mark and stir until the SGZ is dissolved (approximately 3 h). The solution must be stored in a dark bottle in a refrigerator.
5. Tetramethylbenzidine (TMB; $\varepsilon_{655\ nm}$ = 39,000 $M^{-1}cm^{-1}$) (Sigma–Aldrich): the stock is prepared by dissolving 120 mg tetramethylbenzidine base in 50 mL ethanol and adding 2 mL glacial acetic acid with 50 mL of water. Alternatively, a 0.16 M solution in dimethylformamide (DMF) may be prepared. The solution must be stored in a refrigerator.
6. Veratryl alcohol (VA; ε_{310} = 9,000 $M^{-1}cm^{-1}$) (Sigma–Aldrich): Stock solution of VA in toluene (30 mM). The solution must be stored in a refrigerator.

7. Guaiacol (GUA; $\varepsilon_{470}=26,600\ M^{-1}cm^{-1}$) (Sigma–Aldrich): 0.1 M stock solution in phosphate buffer (0.1 M, pH 6.0). The solution must be stored in a refrigerator.
8. Violuric acid (VA) (Sigma–Aldrich): violuric acid is dissolved in DMF as stock solution (0.5 mM). The solution must be stored in a refrigerator.
9. Dyes such as phenol red (PR), methylene blue (MB) and poly-R 478 (Sigma–Aldrich): stock solutions (from 0.05 to 0.2% w/v in phosphate buffer 0.1 M, pH 6.0) are prepared and stored at room temperature in the dark.
10. Tyrosine or preferably tyramine (Sigma–Aldrich): the stock solution (1.0 M) is made in DMF and stored at 4°C.
11. Lignin dimer I, 1-(3,4-dimethoxyphenyl)-2-(2-methoxyphenoxy) propan-1,3-diol (I) prepared according to established procedures (27). The stock solution (32.7 mM) is made in DMF and stored at 4°C.
12. Gallic acid (GA), 3,4-dihydroxybenzaldehyde (DHB), and pyrogallol (PYR) (Sigma–Aldrich): the stock solution (200 μg/mL) is made in methanol and stored at 4°C.
13. Lignin (from Sigma–Aldrich).

2.2. Strains

1. *Escherichia coli* XL1 MRF′ [Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac F′ proAB lacIqZΔM15 Tn10 (Tetr)] from Stratagene – for screening lambda phage expression libraries.
2. *E. coli* XLOLR [Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 thi-1 recA1 gyrA96 relA1 lac F′ proAB lacIqZΔM15 Tn10 (Tetr)] from Stratagene – for screening lambda phage expression libraries.
3. EPI300-T1^R [$^-$mcrA Δ(mrr-hsdRMS-mcrBC) φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ^- rpsL nupG trfA tonA] from Epicentre – for screening pCCFOS fosmid libraries.
4. *E. coli* DH5α [F-ϕ80lacZΔM15 Δ(lacZYA-argF)U169 deoR recA1 endA1 hsdR17(r_k^-, m_k^+) phoA supE44 thi-1 gyrA96 relA1 λ] from Invitrogen – for screening pLAFR3 cosmid libraries.

2.3. Laboratory Equipment

1. 384 Deep-Well Master Block PP, sterile (Greiner Bio-One).
2. 384-Well Master Block PP, sterile (Greiner Bio-One).
3. 96-Well Master Block, PP, 2 mL, sterile (Greiner Bio-One).
4. 96-Well microplates U-bottom, crystal clear, sterile (Greiner Bio-One).

5. 96-Well microplates F-bottom, crystal clear, sterile (Greiner Bio-One).
6. Petri dishes 245 × 245 × 20 mm, sterile (Lab-Center).
7. Microsorb-MV C-18 column (Rainin Instrument Co., Inc.).
8. 96- or 384-pin replicator (Genetix).

3. Methods

3.1. Protein Preparation

1. If possible, use a laccase preparation with known activity (see Note 2). Accurately weigh the amount of the preparation sufficient to obtain laccase activity of 0.7 LAMU (laccase Myceliophthora units)/mL. One LAMU is defined as the amount of enzyme that oxidises one micromol of SGZ per minute under standard conditions (pH 7.5; 30°C). Dilute the test sample on the basis of the anticipated enzyme content to obtain activity between 0.012 and 0.035 LAMU/mL.
2. Protein assay – pour the substrate working solution into a 200 μL (for 96-well plates) or 1 mL (standard cuvettes) 0.05 M sodium citrate/0.1 M phosphate or in 20 mM acetate buffers (pH 4.5–6.5). A standard 40°C assay may be performed. A total of 25 readings are taken for each sample at 5 s intervals. Readings 12–24 are used to calculate the increase of absorbance per minute (ΔAbs/min). The standard working solutions and specific assay conditions are given in step 4–10.
3. Activity calculations – the ΔAbs/min for each well or cuvette containing the test sample into is converted into activity expressed in IU (international units) per milligram. The activity of test samples expressed in U/mg is then calculated using the following formula:

   $$\text{IU/mg} = (A \times \text{Vol} \times D)/W$$

   Where

   A = ΔAbs/min converted to activity (U/mL)
   Vol = volume used to dilute the test sample (mL)
   D = additional dilution of the sample (mL/mL)
   W = weight of the sample (mg)

3.2. Laccase Assays

1. ABTS assay – laccase activity is determined at 420 nm using 0.4 mM ABTS. One enzyme IU is calculated as the formation of one mol of oxidised ABTS per minute using an extinction coefficient of 38,000 $M^{-1}\,cm^{-1}$.
2. DMP, DMA, and MBA assay – the reaction is started by adding the substrate, from a stock solution in methanol, to a final

concentration of 20 μM, and by measuring the absorbance at 468, 310, and 500 nm, in the same order. One enzyme IU is calculated as the formation of one mol of oxidised substrate per minute using an extinction coefficient of 14,800, 9,500, and 38,000 M^{-1} cm^{-1}, respectively.

3. 1-HBT assay – the consumption of this redox mediator is monitored by UV absorption at 230 nm using a final concentration of 1 mM. One enzyme IU is calculated as the formation of one mol of oxidised substrate per minute using an extinction coefficient of 72.8 M^{-1} cm^{-1}.
4. SGZ assay – the reaction mixture consists of 25–75 μM of SGZ, and is followed at 530 nm. One enzyme IU is calculated as the formation of one mol of oxidised substrate per minute using an extinction coefficient of 64,000 M^{-1} cm^{-1}.
5. TMB assay – the assay mixture contains 1.6 mM TMB, and activity is monitored at 655 nm. One enzyme IU is calculated as the formation of one mol of oxidised substrate per minute using an extinction coefficient of 39,000 M^{-1} cm^{-1}.
6. VA assay – the initial rate of oxidation of veratryl alcohol to veratraldehyde is detected by absorption at 310 nm. The assay mixture contains 2.5 mM VA and 0.5 mM H_2O_2. One unit of enzyme activity is considered as the amount of enzyme which oxidises one mol of veratryl alcohol per minute using an extinction coefficient of 9,000 M^{-1} cm^{-1}.
7. GUA assay – the incubation mixture contains from 33 to 4.8 mM GUA and 314 μM H_2O_2. BSA is added because we have noticed in previous experiments that its presence in the assay mixture led to better reproducibility among replicate samples. The reaction can be detected at 650 or 470 nm. One enzyme IU is calculated as the formation of one mol of oxidised substrate per minute using an extinction coefficient of 26,600 M^{-1} cm^{-1}.
8. VA assay – each reaction mixture contains 1 mM VA, and the consumption of this redox mediator is monitored by reversed-phase HPLC by using a Microsorb-MV C-18 column. The mobile phase is acetonitrile–water (35:65) containing 0.1% trifluoroacetic acid. Compounds are detected by UV absorption at 280 nm. A standard violuric acid is used for identification and quantification of the eluted compounds by an external standard analysis.
9. Dyes such as phenol red (PR), methylene blue (MB), and poly-R 478 assay – the reaction mixture for dye decolourisation consists of an aqueous solution of dye, crude laccase and VA or 1-HBT at different concentrations (1, 2, and 5 mM). Dye concentrations are selected to obtain around 1.5 absorbance units at the maximum wavelength in the visible spectrum

(about 40–130 mg/L). All the reaction mixtures are incubated at room temperature, without shaking and in complete darkness. Decolourisation is measured spectrophotometrically from 350 to 750 nm (optimally at 432 nm), by measuring the area under the plot and expressed in terms of percentage.

10. Tyrosine or tyramine assay – the reaction mixture contains 5 mM substrate (tyramine), 50 μM $CuSO_4$, and 5 mM ascorbic acid. The addition of ascorbic acid prevents formation of quinine products, and tyramine was preferred to tyrosine as a substrate owing to its higher V_{max} value compared with tyrosine. The incubation is stopped by acidifying the reaction mixture with 50 μL 0.5 M HCl. After addition of 50 μL of 100 mg/mL $NaNH_2$ and 100 mg/mL $MbNO_3$, samples are allowed to stand for 5 min. Alkalisation with 50 μL of 2 μM NaOH generates a red colour with an A_{max} of circa 490 nm, which is immediately measured. One enzyme IU is calculated as the formation of one mol of oxidised substrate per minute using an extinction coefficient of 3.3 $mM^{-1}cm^{-1}$.
11. Lignin dimer I assay – the oxidation of dimer I and consumption of redox mediators is monitored by reversed-phase HPLC by using a Microsorb-MV C-18 column. The mobile phase is acetonitrile–water (35:65) containing 0.1% trifluoroacetic acid. Compounds are detected by UV absorption at 280 nm. A standard dimer I solution is used for identification and quantification of the eluted compounds by an external standard analysis.
12. GA, DHB and PYR assay – the range of substrate concentrations used in experiments are GA (1–7 mM), DHB (3–9 mM), and PYR (0.6–2 mM). In order to prevent oxidation of substrates by dissolved oxygen, purified oxygen-free nitrogen is passed through substrate solution before each experiment. The changes in absorbance due to oxidation of substrates are monitored at 470 nm for 5 min of incubation. One enzyme IU is calculated as the formation of one mol of oxidised substrate per minute using an extinction coefficient of 6,740 $M^{-1}cm^{-1}$.

3.3. Screening Metagenomes for Laccase Activity

The above substrates may enable detection of laccase or related activities in colonies on agar or crude cell lysates by the production of a fluorophore or chromophore. Assays on agar-plated colonies typically enable the screening of >10^4 variants in a matter of days but are often limited in sensitivity (see Note 3). Soluble products diffuse away from the colony, and hence, only very active variants are detected. This is the case for ABTS agar assay (28). The range of assays that are applicable for crude cell lysates is obviously much wider, but their throughput is rather restricted. In the absence of sophisticated robotics, which is usually unavailable to academic laboratories, only 10^3–10^4 variants are typically

screened (29). These low-to-medium throughput screens have proved effective for the isolation of enzyme variants with improved properties or for the isolation of enzymes from pre-enriched metagenomic libraries as described in a number of reviews (30). However, a far more efficient sampling of large metagenomic libraries is required for the isolation of rare variants or those with dramatically altered phenotypes. Following on from this, the selection-based approach that involves construction of small- to large-insert expression libraries, especially those made in lambda phage, cosmid, or copy-control fosmid vectors, which are further implemented for a direct activity screening (29) is the best option for laccase discovery; however, their scope is rather limited for laccase activity screens. Different implemented assays to be used in agar are described below. For high-throughput liquid screening, the methods below can also be used.

3.3.1. Preparation of Clone Libraries for Screening

Details about the preparation of metagenome libraries is extensively described in the chapter "Molecular Methods to Study Complex Microbial Communities" by Guazzaroni et al. Following the indication provided by Guazzaroni et al., the libraries are prepared, and clones are individually placed in 384-well plates (for cosmid or copy-control fosmid libraries) or as pool of clones or phages (for phage libraries).

3.3.1.1. Expression Lambda Phage Libraries

1. Streak the bacterial glycerol stock (*E. coli* XL1 MRF′) onto the LB agar plates. Incubate the plates overnight at 37°C. Do not add antibiotic to the medium in the following step. The antibiotic will bind to the bacterial cell wall and inhibit the ability of the phage to infect the cell.
2. Inoculate 50 mL of Luria–Bertani (LB) medium, supplemented with 10 mM $MgSO_4$ and 0.2% (w/v) maltose from sterile autoclaved 1 M- and 20%- stock solutions, correspondingly, with a single colony of *E. coli* XL1 MRF′.
3. Grow at 37°C, shaking for 4–6 h (do not grow past an OD_{600} of 1.0). Alternatively, grow overnight at 30°C, with shaking at 200 RPM.
4. Pellet the bacteria at 500 × *g* for 10 min.
5. Gently resuspend the cells in half the original volume with sterile 10 mM $MgSO_4$.
6. Dilute the cells to an OD_{600} of 0.5 with sterile 10 mM $MgSO_4$. The bacteria should preferably be used immediately after dilution.
7. Mix the aliquot of the phage library containing up to a maximum of 10,000 phage particles with 2,000 μL of the cells prepared above with OD_{600} of 0.5 in 50 mL Falcon tube and incubate the tube at 37°C for 15 min gently shaking the tube after each 4 min.

8. Transfer the cells from above step into 45 mL of NZY melted soft agar chilled down to 40–42°C (use the water bath to equilibrate the temperature) containing the substrate solution (see below) and 50 μM $CuSO_4$, and pour onto NZY agar plates (22.5 × 22.5 × 2 cm). Importantly, on the one hand, avoid the solidified agar bits (this will make your screening impossible), and on the other hand, avoid the over-heated agar, which will kill *E. coli*. Wait until the soft agar layer is solidified (keep the open plate under the clean bench) and dry the plate for extra 15 min under clean bench. Incubate the plates overnight at 37°C.

3.3.1.2. pLAFR3 Cosmid and pCCFOS Fosmid Libraries

Replicate manually with the help of a 96- or 384-pin replicator the individual clones onto the LB agar plates containing 10 μg/mL tetracycline (for pLAFR3) or 12.5 μg/mL chloramphenicol (for pCCFOS), the appropriate substrate, and 50 μM $CuSO_4$. Incubate the plates overnight at 37°C. When required, plates should be kept in a dark plate (cover with aluminium paper). Plates (225 × 225× 20 mm) can contain up to 2,304 colonies (replica from six 384-well plates) particles. This protocol may be extended for any kind of libraries, including bacterial chromosomal DNA (BAC) libraries.

3.3.2. Implemented Agar Assays for Metagenome Library Screening

1. ABTS-like detection – the minimal concentration of ABTS to be added into the medium is 3 mM. Positive clones are identified by a light brown colour produced by the oxidation of ABTS to the cationic radical $ABTS^{\bullet+}$. Oxidation of ABTS by laccase results in the production of a green-blue coloured radical cation ($ABTS^{+\bullet}$); however, this colour can be masked in the complex growth media. This method has not been used for library screening, but it has successfully been employed for exploring error-prone PCR libraries in *Saccharomyces cerevisiae* (29).
2. SGZ-like detection – laccase production by library clones is screened by plating hybrid phage-infected cells or clones on appropriate (soft) agar containing 50 μM SGZ. Positive clones were identified by a purple halo, produced by the oxidation of SGZ, on agar plate. This method has successfully been used by Beloqui et al. (20).
3. DMP, DMA, and MBA-like detection – the oxidation products of these substrates cannot be detected in the visible spectrum; however, if reacting with iodide, a triodide that has an intense yellow colour is formed. Therefore, the presence of the yellow colour in the agar plate is determined by adding the solution of 50 μM substrate in 0.4% agarose or agar made in 0.05 M sodium citrate/0.1 M phosphate or in 20 mM acetate buffers (pH 4.5–6.5), containing 0.1 mM iodide over

the agar plate with bacterial or phage colonies pre-grown overnight. The solution must be well shaken and poured over and allowed to solidify. The plates must be held at the room temperature for 30 min. A yellow colour indicates the reaction zone.

4. TMB-like detection – plates with bacterial or phage colonies are incubated for 12 h at 37°C and then are covered with a second layer containing the substrate (TMB to a final concentration of 2 mM in a 20-mL 0.1 M potassium phosphate buffer, pH 5.4, 0.4% (w/v) agarose, plus 200 μM of H_2O_2 to initiate the reaction). Positive clones form a brown precipitate. A blue-coloured charge-transfer complex is generated.
5. GUA-like detection – plates with bacterial or phage colonies are incubated for 12 h at 37°C and then are covered with a second layer containing the substrate (GUA to a final concentration of 5 mM in a 20-mL 0.1 M potassium phosphate buffer, pH 5.4, 0.4% (w/v) agarose, plus 200 μM of H_2O_2 to initiate the reaction). GUA oxidation reactions are indicated by an amber-coloured product around active clones.
6. VA-like detection – VA is added to the medium at a final concentration of 1.6 mM together with 20 μM H_2O_2, and plates are incubated for 12 h at 37°C. Active colonies are detected by exposition to UV light (280 nm).
7. PR, MB, and poly-R 478-like detection – the dye decolourization consists of the following mixture: 0.025% w/v dye is added to the agar medium, and plates are incubated for 12 h at 37°C; laccase activity is detected by monitoring the disappearance of the substrate.
8. Tyrosine or tyramine-like detection – tyrosine or tyramine are added to the medium at a final concentration of 100 mM, and plates are incubated for 12 h at 37°C. Substrate oxidation, and thereby active clones, is detected by exposition to UV light (280 nm).
9. Lignin-like detection – lignin is treated with ten volumes of H_3PO_4 1 M, after which the products are washed through a filter with 100 volumes H_2O and further dried at 70°C overnight. The resulting product is added to the agar medium to a final concentration of 0.05% w/v together with 20 μM H_2O_2, and plates are incubated for 12 h at 37°C. The oxidation of lignin through the laccase activity is detected by monitoring the disappearance of the strong brown colour of the substrate.
10. GA, DHB, and PYR-like detection – positive clones are detected using similar protocol as described for DMP, DMA, and MBA (see above).

4. Conclusions

Laccases have a great interest for both basic structural–functional research on proteins and applied science. Two main reasons may account for this. First, their metal content, redox potential, and cluster-like structure exemplify the structural complexity of oxidoreductases' protein family. Second, they use molecular oxygen to oxidise various aromatic and non-aromatic compounds which is very relevant for the bioremediation and energy production processes. Although, they have been extensively studied in pure cultures of fungi, the recent identification of laccase activity in bacteria and also in metagenome may indicate that the mining of the vast genetic diversity by metagenomic approaches is a promising strategy for the identification and isolation of novel laccases. Identification of such enzymes will enhance our knowledge of the structure, function, and evolution of enzymes and will allow definition of many new enzyme families. The application of high-throughput screening approaches to the genomic and metagenomic libraries will enable us to identify novel enzymes for a wider scope of characteristics. These newly identified enzymes can serve as the ideal starting point for the directed evolution of novel laccases with improved properties.

5. Notes

1. Laccase substrates are best stored at 4°C in appropriate solvent. Store them in small opaque bottles as it may decline in quality after preparation.
2. Unless stated otherwise, all laccase containing solutions should be prepared in 10 mM phosphate buffer (0.1 M, pH 6.0) or 10 mM sodium acetate buffer (pH 4.5) and stored at 4°C. We have found that this pH is optimal for enzyme activity and stability.
3. The screening protocols in agar or liquid medium when screening a library of clones can be adapted for many substrates. For substrates sensitive to light and auto-degradation, we recommend incorporating the substrate after overnight cultivation.

Acknowledgements

This research was supported by the Spanish MEC BIO2006-11738, CSD2007-00005 and GEN2006-27750-C-4-E projects. A.B thanks the Spanish MEC for a FPU fellowship. P.N.G. was

supported by Grant 0313751K from the Federal Ministry for Science and Education (BMBF) within the GenoMik-Plus initiative.

References

1. Thurston, C.F. (1994) The structure and function of fungal laccases. *Microbiology* 140, 19–26.
2. Sariaslani, F.S. (1989) Microbial enzymes for oxidation of organic molecules. *Crit Rev Biotechnol* 9, 171–257.
3. Bourbonnais, R. and Paice, M. (1990) Oxidation of non-phenolic substrates. An expanded role for laccase in lignin biodegradation. *FEBS Lett* 267, 99–102.
4. Eggert, C., Temp, U., Dean, J.F., and Eriksson, K.E. (1996) A fungal metabolite mediates degradation of non-phenolic lignin structures and synthetic lignin by laccase. *FEBS Lett* 391, 144–148.
5. Solomon, E.I., Sundaram, U.M., and Machonkin, T.E. (1996) Multicopper oxidases and oxygenases. *Chem Rev* 96, 2563–2605.
6. Camarero, S., Ibarra, D., Martínez, M.J., and Martínez, A.T. (2005) Lignin-derived compounds as efficient laccase mediators for decolorization of different types of recalcitrant dyes. *Appl Environ Microbiol* 71, 1775–1784.
7. Durán, N., Rosa, M.A., D'Annibale, A., and Gianfreda, L. (2002) Applications of laccases and tyrosinases (phenoloxidases) immobilized on different supports: a review. *Enzyme Microb Tech* 31, 907–931.
8. Wesenberg, D., Kyriakides, I., and Agathos, S.N. (2003) White-rot fungi and their enzymes for the treatment of industrial dye effluents. *Biotechnol Adv* 22, 161–187.
9. Arias, M.E., Arenas, M., Rodríguez, J., Soliveri, J., Ball, A.S., and Hernández, M. (2003) Kraft pulp biobleaching and mediated oxidation of a nonphenolic substrate by laccase from *Streptomyces cyaneus* CECT 3335. *Appl Environ Microbiol* 69, 1953–1958.
10. Bertrand, T., Jolivalt, C., Btiozzo, P., Caminade, E., Joly, N., Madzak, C., et al. (2002) Crystal structure of a four-copper laccase complexed with an arylamine: insights into substrate recognition and correlation with kinetics. *Biochemistry* 41, 7325–7333.
11. Gianfreda, L., Xu, F., and Bollag, J. (1999) Laccases: a useful group of oxidoreductive enzymes. *Bioremediat J* 3, 1–25.
12. Huttermann, A., Mai, C., and Kharazipour, A. (2001) Modification of lignin for the production of new compounded materials. *Appl Microbiol Biotechnol* 55, 387–394.
13. Mayer, A.M. and Staples, R.C. (2002) Laccase: new functions for an old enzyme. *Phytochemistry* 60, 551–565.
14. Parkinson, N.M., Conyers, C.M., Keen, J.N., MacNicoll, A.D., Smith, I., and Weaver, R.J. (2003) cDNAs encoding large venom proteins from the parasitoid *Pimpla hypochondriaca* was identified by random sequence analysis. *Comp Biochem Physiol C-Toxicol Pharmacol* 134, 513–520.
15. Enguita, F.J., Martins, L.O., Henriques, A.O., and Carrondo, M.A. (2003) Crystal structure of a bacterial endospore coat component – A laccase with enhanced thermostability properties. *J Biol Chem* 278, 19416–19425.
16. Alexandre, G. and Zhulin, I.B. (2000) Laccases are widspread in bacteria. *Trends Biotechnol* 18, 41–42.
17. Claus, H. (2003) Laccases and their occurrence in prokaryotes. *Arch Microbiol* 179, 145–150.
18. Claus, H. (2004) Laccases: structure, reactions, distribution. *Micron* 35, 93–96.
19. Vieites, J.M., Guazzaroni, M.E., Beloqui, A., Golyshin, P.N., and Ferrer, M. (2009) Metagenomics approaches in systems microbiology. *FEMS Microbiol Rev* 33, 236–255.
20. Beloqui, A., Pita, M., Yakimov, M.M., Polaina, J., Zumárraga, M., García-Arellano, H., et al. (2005) Novel polyphenol oxidase mined from a metagenome expresion library of bovine rumen: biochemical properties, structural analysis and phylogenetic relationships. *J Biol Chem* 281, 22933–22942.
21. Martins, L.O., Soares, C.M., Pereira, M.M., Teixeira, M., Costa, T., Jones, G.H., et al. (2002) Molecular and biochemical characterization of a highly stable bacterial laccase that occurs as a structural component of the *Bacillus subtilis* endospore coat. *J Biol Chem* 277, 18849–18859.
22. Ruijssenaars, H.J. and Hartmans, S. (2004) A cloned *Bacillus halodurans* multicopper oxidase exhibiting alkaline laccase activity. *Appl Microbiol Biotechnol* 65, 177–182.

23. Ducros, V., Brzozowski, A. M., Wilson, K. S., Brown, S. H., Ostergaard, P., Schneider, P., et al. (1998) Crystal structure of the type-2 Cu depleted laccase from *Coprinus cinereus* at 2.2 Å resolution. *Nature Struct Biol* 5, 310–316.

24. Antorini, M., Herpoel-Gimbert, I., Choinowski, T., Sigoillot, J.C., Asther, M., Winterhalter, K., et al. (2002) Purification, crystallisation and X-ray diffraction study of fully functional laccases from two ligninolytic fungi. *Biochim Biophys Acta* 1594, 109–114.

25. Hakulinen, N., Kiiskinen, L.L., Kruus, K., Saloheimo, M., Paananen, A., Koivula, A., et al. (2002) Crystal structure of a laccase from *Melanocarpus albomyces* with an intact trinuclear copper site. *Nature Struct Biol* 9, 601–605.

26. McGuirl, M.A. and Dooley, D.M. (1999). Copper-containing oxidases. *Curr Opin Chem Biol* 3, 138–144.

27. Adler, E. and Eriksson, E. (1955) Guaiacylglycerol and its β-guaiacyl ether. *Acta Chem Scand* 9, 341–345.

28. Zumárraga, M., Bulter, T., Shleev, S., Polaina, J., Martínez-Arias, A., Plou, et al. (2007) In vitro evolution of a fungal laccase in high concentrations of organic cosolvents. *Chem Biol* 14, 1052–1064.

29. Ferrer, M., Beloqui, A., Vieites, J.M., Guazzaroni, M.E., Berger, I., and Aaroni, A. (2009) Interplay of metagenomics and *in vitro* compartmentalization. *Microb Biotechmol* 2, 31–39.

30. Ferrer, M., Beloqui, A., Timmis, K. N., and Golyshin P. N. (2009) Metagenomics for mining new genetic resources of microbial communities. *J Mol Microbiol Biotechnol* 16, 109–123.

Chapter 14

Screening for *N*-AHSL-Based-Signaling Interfering Enzymes

Phil M. Oger and Stéphane Uroz

Abstract

Quorum sensing (QS)-based signaling is a widespread pathway used by bacteria for the regulation of functions involved in relation to their environment or host. QS relies upon the production, accumulation, and perception of small diffusible molecules by the bacterial population, hence linking high gene expression with high cell population densities. Amongst the different QS signal molecules, an important class of signal molecules is the *N*-acyl homoserine lactone (*N*-AHSL) class. In pathogens such as *Erwinia* or *Pseudomonas*, *N*-AHSL-based QS is crucial to overcome the host defenses and ensure a successful infection. Interfering with QS regulation allows the alga *Delisea pulchra* to avoid surface colonization by bacteria. Thus, interfering in the QS regulation of pathogenic bacteria is a promising antibiotic-free antibacterial therapeutic strategy. To date, two *N*-AHSL lactonase and one amidohydrolase families of *N*-ASHL degradation enzymes have been characterized and proven to be efficient in vitro to control *N*-AHSL-based QS-regulated functions in pathogens.

Key words: *N*-acyl homoserine lactone, Quorum sensing, Quorum quenching, *N*-AHSL lactonase, *N*-AHSL acylase, *N*-AHSL amidohydrolase

1. Introduction

Gram-negative bacteria couple gene expression to population density by a regulatory mechanism named *q*uorum *s*ensing (QS). QS relies upon the production and the perception of one or more signal molecules by the bacterial population (1, 2). An important class of these signals is the *N*-acyl *h*omo*s*erine *l*actone (*N*-AHSL) class. QS regulates pathogenicity or pathogenicity-related functions in bacteria of medical or environmental importance such as the human pathogen *Pseudomonas aeruginosa* or the plant pathogens *Erwinia carotovora* and *Agrobacterium tumefaciens* (3, 4). If QS is an important component of the adaptation strategy of bacteria to their environment, one might suspect that competing

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668, DOI 10.1007/978-1-60761-823-2_14, © Springer Science+Business Media, LLC 2010

bacteria/eukarya might have developed strategies to interfere with this communication system. Indeed, QS interference was reported through the production of antagonists or the production of *N*-AHSL degradation enzymes (*N*-AHSLases) in various organisms from human, plant, and fungi to bacteria (5, 6). Whatever the physiological role of the *N*-AHSLases in their host, they have been used to interfere efficiently with the expression of QS-regulated functions in bacteria (7). Thus, interfering with QS regulation, a strategy, coined the term quorum quenching appears as one of the promising nonantibiotic-based therapeutic strategies for the future (8).

N-AHSLs exhibit a conserved structure, with a backbone composed of a lactone ring derived from the lactonization of homoserine, *N*-linked to an acyl chain via an amide bond (Fig. 1). Variation in *N*-acyl chain length and the oxidation status of *N*-AHSLs provide for specificity of the signal. Four chemical or enzymatic alterations of the structure are known to occur (Fig. 1) two of which, lactonolysis and amidohydrolysis, generate QS inactive molecules. Amidohydrolysis cleaves the *N*-AHSL molecule irreversibly into two QS inactive molecules, homoserine lactone (HSL) and the corresponding acyl chain. On the contrary, lactonolysis is a reversible reaction opening the lactone ring of the HSL moiety to yield *N*-acyl homoserine (*N*-AHS). It occurs spontaneously under

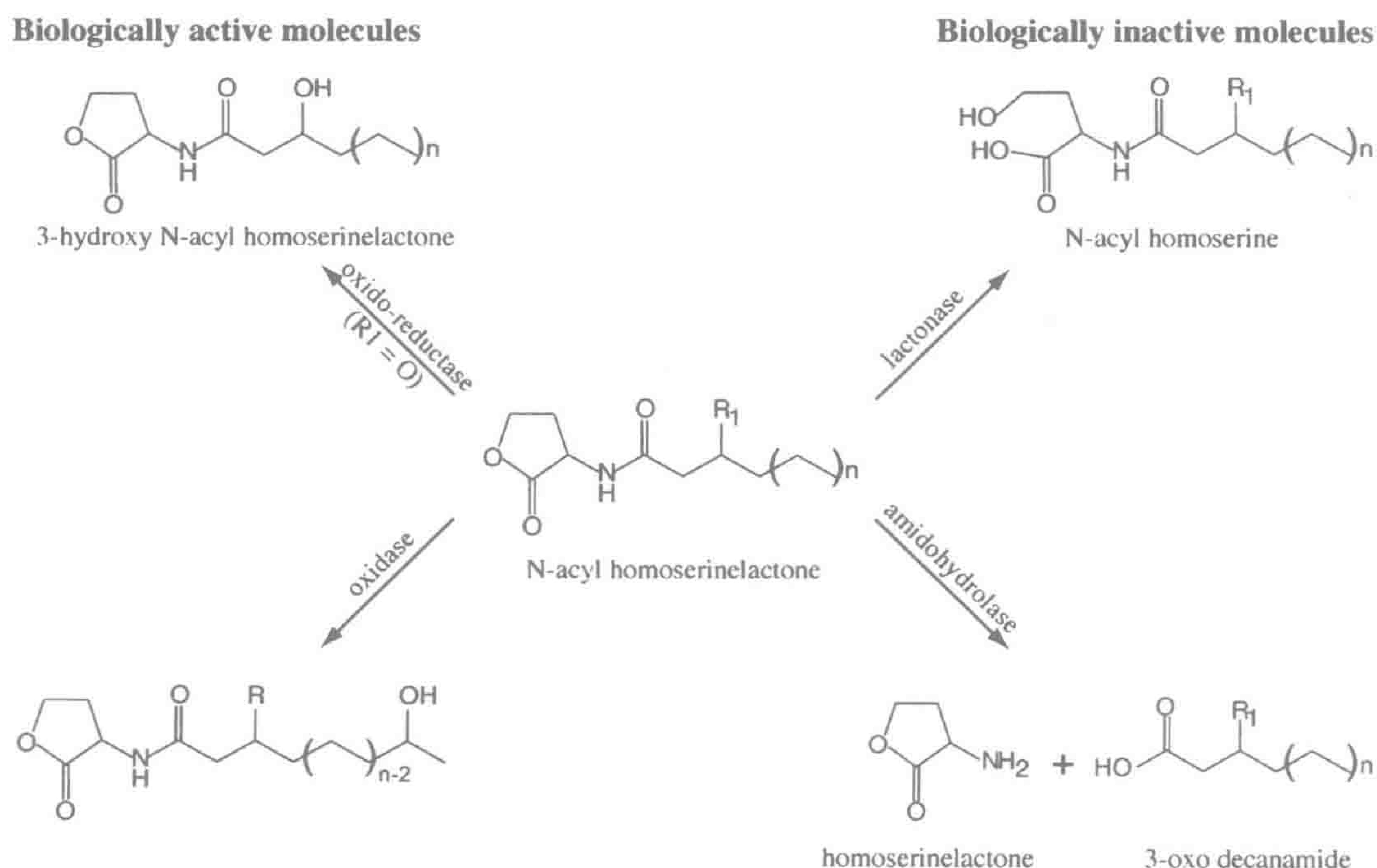


Fig. 1. *N*-AHSL chemical and ezymatic alterations. *Center*: common structure of *N*-AHSLs (R1 = OH or O; $0 \leq n \leq 6$). *Left*: biologically active derivatives of *N*-AHSL following oxidase and oxidoreductase attacks. *Right*: biologically inactive *N*-AHSL derivatives following lactonase or amidohydrolase degradation.

basic pH, while low pH favors the recircularization of the lactone (9). Despite the large diversity of *N*-AHSL-degrading organisms identified to date, only three families of *N*-AHSL-inactivating enzymes have been described: the AiiA and QsdA *N*-AHSL lactonase families (7, 10) and the AiiD (11) *N*-AHSL amidohydrolase (or acylase) family. Since they irreversibly cleave the signal molecule, *N*-AHSL amidohydrolases have a greater biotechnological potential than *N*-AHSL lactonases (see Note 1). A short procedure, applicable from wild-type environmental isolates to purified proteins alike, allows to quickly screen for and characterize these enzymes.

2. Materials

2.1. Strains and Growth Media for Cell Cultures

1. *N*-AHSL sensor systems (see Note 2): Sensor system for 3-oxo and 3-hydroxy *N*-AHSL (3O, and 3OH, *N*-AHSL, respectively) *A. tumefaciens* strain NTL4(pZLR4) (12). This strain should be maintained and cultured on gentamycin 100 mg/L. Sensor system for short-chain *N*-AHSLs *Chromobacterium violaceum* strain CV026 (13). This strain should be cultured in Luria Broth with 5 g NaCl per liter. It cannot be maintained for long periods on plates, and should be streaked regularly from frozen stocks.
2. Low-salt Luria Broth (5 g NaCl/L, Gibco). When necessary, this medium is buffered to pH 6.5 with 100 mM phosphate buffer to avoid spontaneous degradation of *N*-AHSLs. To prepare 1 L of pH 6.5-buffered LB, resuspend 20 g of LB powder into 900 mL of water, then add 27.8 mL of 1 M K_2HPO_4 and 72.2 mL of 1 M KH_2PO_4. Sterilize by autoclaving.
3. AB minimal medium is prepared from stock solutions of 20× AB salts and 20× AB buffer and sterile water for liquid media and sterile water agar for solid media.
   (a) 20× AB salts (per liter): 20 g NH_4Cl, 6 g $MgSO_4{\cdot}7H_2O$, 3 g KCl, 200 mg $CaCl_2$, 50 mg $FeSO_4{\cdot}7H_2O$; sterilize by autoclaving.
   (b) 20× AB buffer (per liter): 60 g K_2HPO_4, 23 g NaH_2PO_4; adjust the pH to 7 if necessary; Sterilize by autoclaving.

5 mM mannitol from a stock solution at 100 mM is added as a carbon source. When necessary, gentamycin (100 mg/L) and X-gal (40 mg/L) are added to the medium.

4. Phosphate-buffered saline solution (PBS 1×): 8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, 0.24 g KH_2PO_4; Dissolve in 800 mL of distilled H_2O. Adjust the pH to 6.5 with HCl. Add H_2O to 1 L. Sterilize by autoclaving.

2.2. N-AHSL Degradation Assays

1. Transilluminator, 315 nm.
2. *N*-AHSL solutions in ethyl acetate (1 mM and 10 μM). Most *N*-AHSLs can be purchased from Sigma–Aldrich (Sigma–Aldrich, Saint-Louis, MO). The others can be purchased from Pr. Paul Williams (Nottingham University, UK).
3. Dansyl chloride (3.7 M in acetone stock solution); HCl 5 and 0.2 M; HPLC-grade dichloromethane; acetonitrile; ethyl acetate (Sigma–Aldrich, Saint-Louis, MO).
4. Bradford kit for protein quantification (Sigma–Aldrich, Saint-Louis, MO).

2.3. Thin-Layer Chromatography

1. Whatman 3 mM filter paper (Whatman, Springfield Mill, UK).
2. Glass TLC Developing Tank for 20 cm × 20 cm TLC plates (Whatman, Springfield Mill, UK).
3. Glass C18 coated TLC plates with 200 μM coating. We use Partisil® KC18 TLC plates, Silica gel 60 Å (Whatman, Springfield Mill, UK).
4. Methanol, analytical grade (Sigma–Aldrich, Saint-Louis, MO).
5. Overlay preparation: Sterilize by autoclaving 88 mL of soft water agar (7 g/L), then add to the medium 5 mL of each 20× AB salts and 20× AB buffer and 2 mL of 100 mM mannitol solution. Cool until it reaches ~50–55°C, and then add 150 μL of X-gal (40 mg/mL).
6. Custom made TLC overlaying container (see Note 3). This container is made out of 5 mm-thick plexiglass. The base of the container is a 25 cm wide square, in which four 3 cm-wide holes placed approximately 5 cm from each corner along the diagonals have been drilled. These holes allow the user to access the plate from underneath and push to release it after the overlay has been solidified. Two centimeter-wide bars glued on top of the base form a 20.2 cm × 20.2 cm 5 mm deep internal space which will accept the TLC plate. It is important to allow some extra spacing around the TLC to facilitate the extraction of the plate after solidification of the agar. The thickness of the overlay is 3 MM.

2.4. HPLC

1. Waters 625 HPLC system (Waters Corp., Milford, MA) coupled with a Waters 996 PDA photodiode array detector (operating with a Millennium 2010 Chromatography Manager) equipped with a Kromasil C8 5 μM column, 2.1 × 250 mm (Jones Chromatography, Mid Glamorgan, UK) or equivalent for the identification of amidohydrolysis degradation products.
2. Waters HPLC system equipped with a Waters separation module 2659 coupled to a Waters Micromass ZQ200

electrospray ionization–mass spectrometry detector and an Atlantis T3 reverse-phase column, 4.6 × 150 mm (Waters Corp., Milford, MA) for the detection of lactonolysis identification products.

3. Methods

Since the discovery of the QS regulation system, several very sensitive sensor strains have been designed for *N*-AHSL detection. These are based on the same principle: the gene responsible for the synthesis of the *N*-AHSL has been mutated and can only respond to the exogenous *N*-AHSLs. Reporter genes can be either native, such as the production of the violacein pigment in the *Chromobacterium* sensor system (13), or engineered, such as to produce β-galactosidase in the *Agrobacterium* sensor system (12).

The screen for *N*-AHSL degradation enzymes proceeds in four steps. First, bacterial strains or clones are screened for their ability to inhibit one of the QS sensor systems (see Subheading 3.1). This first step is not specific for *N*-AHSL degradation enzymes, but allows to screen the molecules/activities that may inhibit the detection of the *N*-AHSL by the sensor, including molecules that might interfere with its growth. Thus, the ability of each clone/strain to degrade *N*-AHSLs or inhibit their detection is determined in a second step (see Subheading 3.3). The third step allows to differentiate between lactonases and amidohydrolases (see Subheading 3.3). The fourth step is a confirmation of this differentiation in which the degradation products of the *N*-AHSLs are characterized (see Subheading 3.4).

3.1. Microplate Fast Screening of N-AHSL Degradation Active Clones (14)

1. Grow the bacterial clones in 200 μL of LB supplemented with the appropriate antibiotics in microtiter plates 24 h at 30°C (or 37°C for *E. coli*) (see Note 4).
2. Subculture the clones into 200 μL of fresh pH 6.5-buffered LB medium without antibiotics but supplemented with 25 μM of the appropriate *N*-AHSL. Incubate for up to 2 days at 25°C (see Note 4). Since *N*-AHSL may be spontaneously degraded in buffered LB medium over long incubation period, it is important to include a spontaneous degradation control. It consists of a noninoculated growth medium supplemented with the same amount of *N*-AHSL.
3. Transfer 5 μL aliquots of the bacterial suspensions to a 96-well microtiter plate containing 200 μL of solidified, pH 6.5-buffered LB agar (16 g/L) medium. Kill the bacteria by UV irradiation

by placing the microtiter plates upside down on a transilluminator for 10 min.

4. Overlay the wells with 10 μL of an overnight culture of the reporter strain *C. violaceum* CV026.
5. Monitor violacein (purple pigment) production after 24 h of incubation at 28°C.
6. Wells in which no violacein production occurs are indicative of putative positive *N*-AHSL degrading clones/strains (Fig. 2). ATTENTION: The lack of violacein production may also reflect other activities due to molecules inhibiting the growth of the sensor, or inhibiting the recognition of the *N*-AHSLs by the sensor. Thus, the ability of the positive clones to effectively degrade the *N*-AHSLs needs to be confirmed by separating the degradation products by TLC (see Subheading 3.4).

3.2. Preparation of Resting Cells (RC) and Cell Crude Extracts (CCE) for N-AHSL Degradation Assays (15)

3.2.1. Preparation of RC

1. Grow cells in LB until late exponential phase.
2. Pellet the cells by centrifugation. Resuspend the cells in PBS. Adjust the cell concentration to 10^9 cells/mL by measuring the OD at 600 nm.
3. Wash cells twice in one-tenth volume of PBS buffer (pH 6.5).
4. Resuspend the cells in one-tenth volume of PBS buffer (pH 6.5). Cell concentration of RC is 10^{10} cells/mL.

3.2.2. Preparation of CCE

1. Cycle the RC suspension five times in a cell disrupter (Constant Systems Cell Disrupter) under 15 kPa pressure.

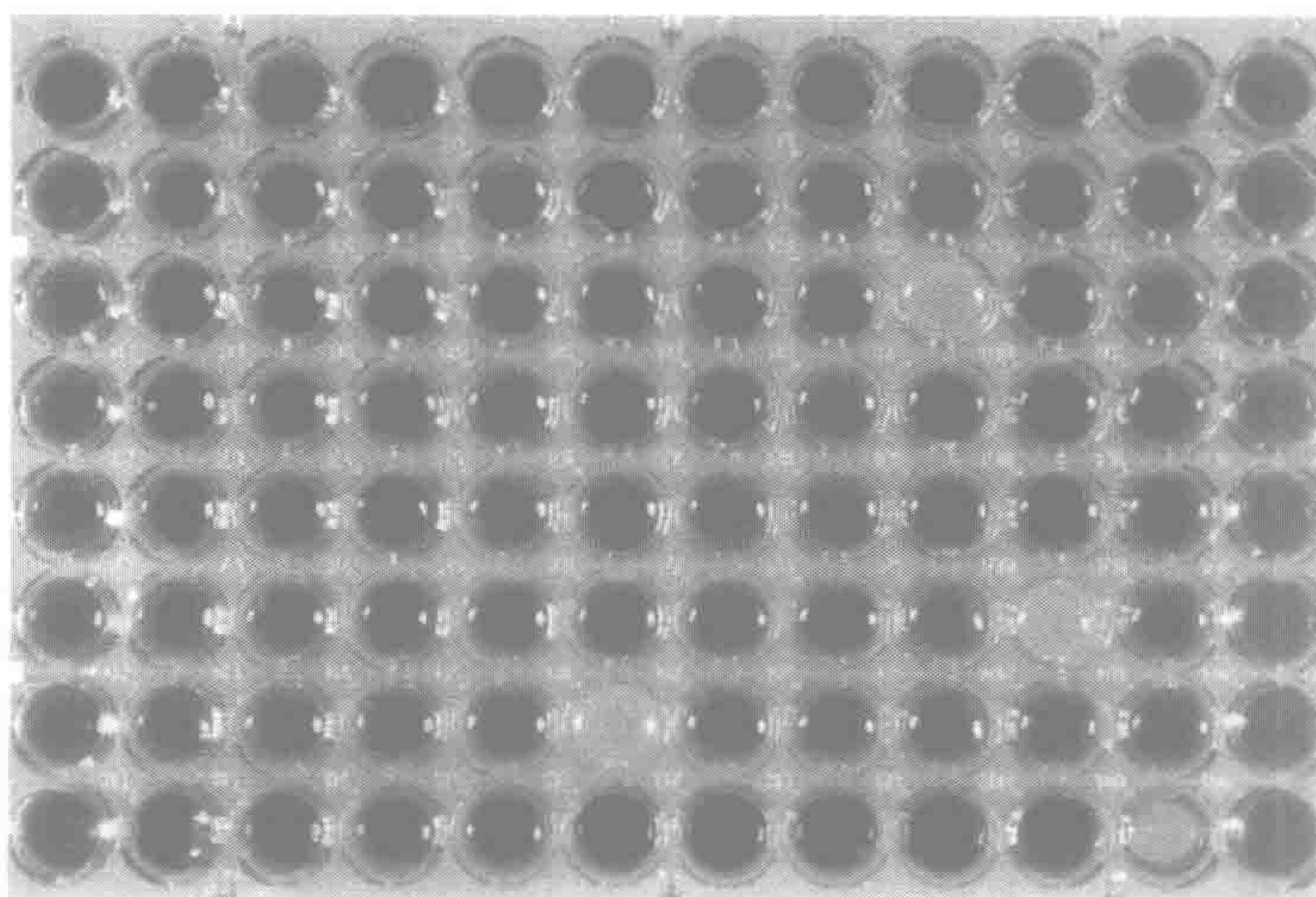


Fig. 2. *N*-AHSL degradation microplate assay of *E. coli* clones expressing a genomic bank of the *Rhodococcus erythropolis* strain W2 imaged after 24 h of incubation. Colorless wells indicate *N*-AHSL degradation.

2. Remove cell debris by centrifugation (120 min, 4°C, 10,000 × *g*).
3. Filter the supernatant through a 0.22 μM membrane.
4. Adjust the protein concentration to 0.5 mg/mL using the Bradford Protein Quantification method and store at 4°C.

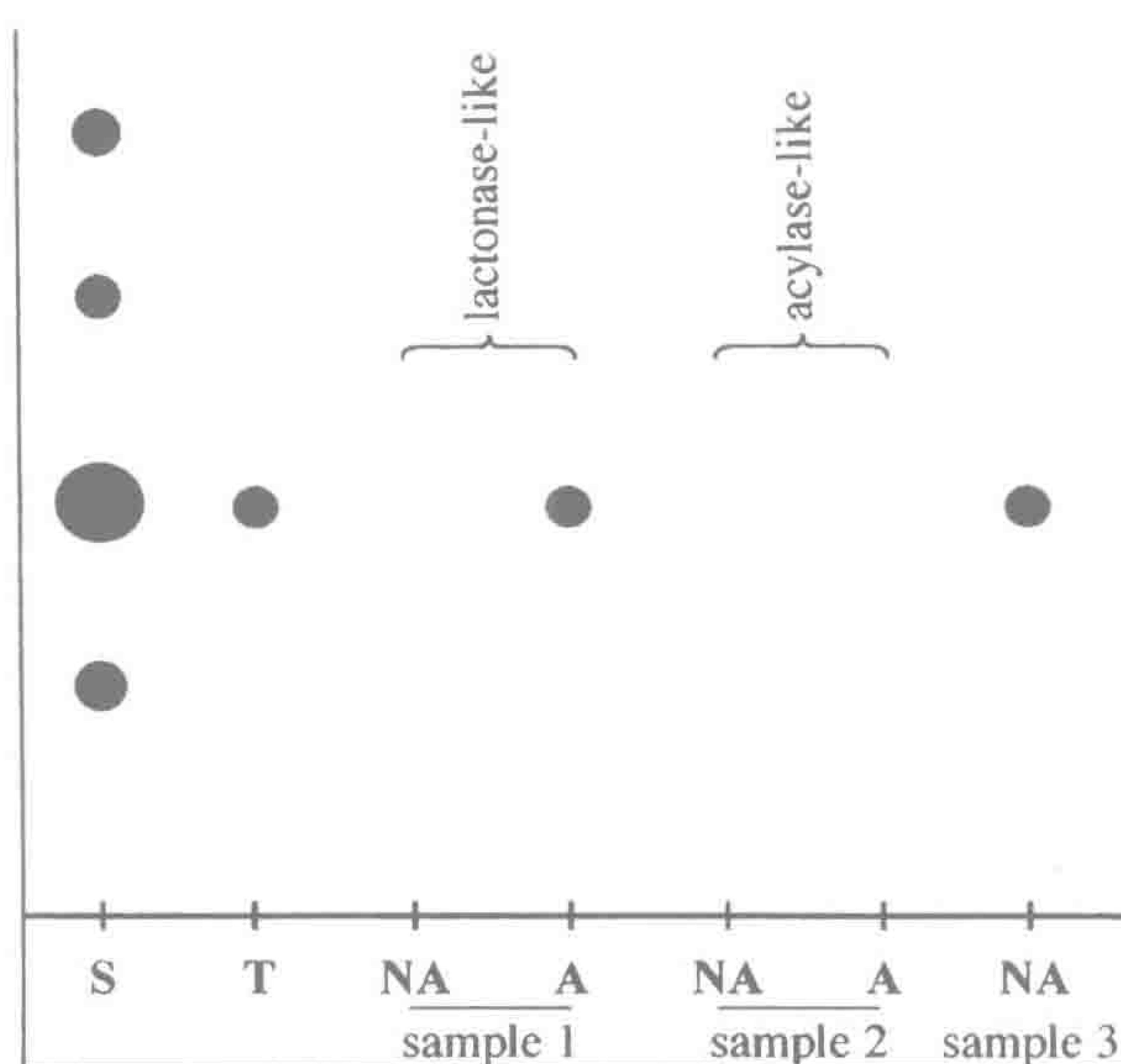
3.3. N-AHSL Lactonase and Acylase Activity Screen/N-AHSL Degradation Confirmation Test

Positive wells in the microplate assay group bacterial strains or clones capable of *N*-AHSL degradation as well as strains/clones with sensor interfering abilities. To detect the fraction of *N*-AHSL degraders, *N*-AHSLs and putative inhibitory molecules present in the growth medium are separated by thin-layer chromatography and detected using the QS sensor. Only clones with *N*-AHSL degradation abilities will fail to induce the QS sensor in both assays (For the detection of false-positive clones, proceed directly to step 1 of subheader 3.3.2, spotting the supernatant obtained in step 2 from subheader 3.1).

The same approach is used to differentiate clones harboring lactonase and amidohydrolase activities. Lactonolysis of *N*-AHSLs yields *N*-acyl homoserine (Fig. 1). This reaction is reversible under low pH, and the *N*-AHSL molecule can thus be regenerated (16). On the contrary, the amidohydrolysis is irreversible. This divergence is exploited in a test to quickly differentiate lactonases from acylases in which one runs side by side on a TLC plate the products of an *N*-AHSL degradation reaction and a subsample acidified to induce lactonization (Fig. 3).

3.3.1. Preparation of the Samples (15)

1. Into a clean 2-mL microcentrifuge tube, add 50 μL of a 10 μM *N*-AHSL stock solution. Evaporate to dryness (see Notes 5 and 6).
2. Add 500 μL of RC to the tube. Vortex for 1 min to dissolve the *N*-AHSL (see Note 7).
3. Incubate for up to 6 h at 25°C (see Note 8).
4. Centrifuge the tube at full speed to pellet the cells. Transfer the supernatant into two clean 2-mL microcentrifuge tubes (250 μL each).
5. In the first tube, add one volume of ethyl acetate to stop and extract the reaction. Vortex for 1 min. Allow the aqueous and ethyl acetate phases to separate for 10 min or centrifuge for 1 min. Transfer the upper phase to a clean tube and evaporate to dryness.
6. To the second tube, add 5 M HCl to acidify the medium to pH 2.
7. Incubate for 24 h at 4°C.
8. Stop and extract the reaction as indicated in step 5.
9. Dissolve the residues in 100 μL of ethyl acetate (see Note 9).



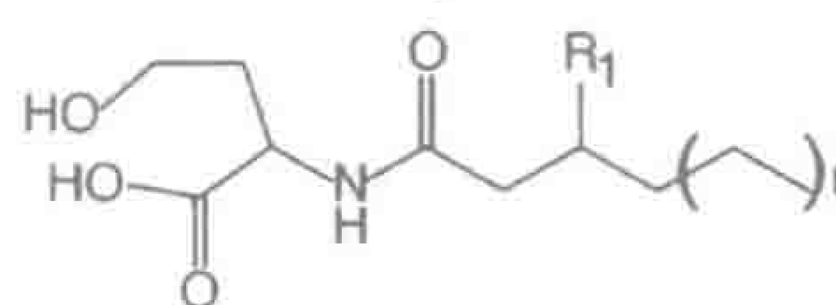
Separation by TLC
Detection with N-AHSL biosensor

non acidified sample

acidified sample

degradation mixture

degradation reaction



N-acyl homoserine

Fig. 3. *N*-AHSL lactonase/acylase differentiation scheme. For each reaction, one sample is acidified (A) and the remnant is not (NA). Sample 1 and 2 present a sketch of results obtained for a lactonase and amidohydrolase, respectively. **S**: *N*-AHSL set of synthetic standards. Sample 3 presents a sketch of results obtained for a false-positive clone in the microplate assay. **T**: positive control, undigested *N*-AHSL.

3.3.2. Preparation, Development, and Revelation of the TLC Plates (17)

1. These instructions assume the use of 20 × 20 cm glass TLC plates, an *Agrobacterium*-based *N*-AHSL detection system, and a custom-made TLC overlaying container (see Note 10).
2. Transfer a single colony of the bacterial sensor strain into 5 mL of AB medium supplemented with 5 mM mannitol and gentamycin (100 μg/mL). Incubate overnight at 30°C under vigorous shaking.
3. The next morning, transfer the 5 mL preculture into 45 mL of the same medium. Incubate at 30°C until late exponential phase, ca. 6 h.
4. Mark the spotting line on a clean TLC plate with a pencil. Care should be taken during the manipulations to avoid dropping organics accidentally onto the TLC plates. Samples should be spotted 2 cm from the bottom of the plate, 2 cm apart from each other. Mark a line 15 cm above from depot line as a guide to know when to stop the chromatography.
5. Spot 1 μL of each samples and standards. Standards should comprise at the least the original *N*-AHSL (see Note 11).
6. Wait until TLC plate is dry.
7. Fill the Glass TLC developing tank with 200 mL of running solution (methanol:water, 60:40, v:v).
8. Cover the inside of the glass TLC developing tank with running solution saturated Whatman 3 MM paper. This step is important to get a linear running front in large TLC developing tanks.
9. Run the plate until it reaches the top line, for approximately 2 h.
10. Take the plate out and dry in a fume hood for 10 min.
11. Mix the reporter strain culture (50 mL) with the cooled overlay medium (100 mL) by shaking gently to avoid the formation of bubbles.
12. Place the plate in the custom-made overlaying container and gently pour the overlay on the plate. Remove the excess of medium and bubbles by running a plastic ruler over the container (see Note 12).
13. Wait until the soft agar gets solidified.
14. Loosen the medium from the sides of the container with a flat spatula to take the plate out of the container.
15. Place the overlaid TLC plate in a plastic container with a paper towel at the bottom (see Note 13). The paper towel is used to help take out the plate out of container after incubation.
16. Incubate the plate with lid closed overnight at 28°C.

17. The plate should show blue dots according to the standard used. Plates are ready to be imaged if the color is sufficiently developed. The drying of the plate greatly improves the contrast but may not be necessary if only presence or absence of a given spot is needed. It is, however, convenient for the storage of the revealed TLC. To dry the plates proceed as follows.
18. Remove the plate from the plastic container and place in the back of a fume hood to dry. Take the plate out of the hood when the plate is close to be completely dried. Overdrying the plate results in the curling and cracking of the C18-layer. Let the plate sit at room temperature to completely dry slowly.

3.3.3. Interpretation of TLC plates (see Note 14)

1. The presence of a lactonase activity will be evidenced by the presence of a blue spot in the acidified sample lane with an Rf identical to, e.g., migrating at the same distance as, the starting *N*-AHSL (Fig. 3, sample 1, lane A), concomitant with the absence of a spot in the nonacidified sample (Fig. 3, sample 1 lane NA).
2. The absence of spots in both acidified and nonacidified lanes is evidence for a degradation activity not involving a lactonase, e.g., to date indication of an amydohydrolase activity (Fig. 3, sample 2).
3. False-positive clones for the degradation of *N*-AHSLs will be evidenced after TLC separation by the presence of a spot with an Rf identical to the starting *N*-AHSL in the NA lane (sample 3).

3.4. Identification of N-AHSL Lactonase Activities by HPLC-MS

1. *N*-AHSL degradation reactions are set as described above for the TLC plate assay, except that one should use 50 μL of a 1 mM *N*-AHSL solution to stop the complete reaction after the appropriate incubation time and dissolve the reaction into 50 μL of ethyl acetate (see Note 7).
2. Inject 10 μL of the reaction mixture into the HPLC system.
3. Elution: water/formic acid 0.1% (solvent A) and acetonitrile/formic acid 0.1% (solvent B) under the following elution sequence: 100% A 5 min; linear gradient 100% A 0% B to reach 80% A and 20% B 5 min; 80% A and 20% B 10 min. Between two samples, the column is rinsed by applying a linear gradient to reach 100% B (2 min), and 100% B (3 min). Column is then reequilibrated with 100% A for 7 min at a flow rate of 2 mL/ min (see Note 15).
4. Under our experimental conditions, the C6-HS and C6-HSL harbors retention times of 15.81 and 21.00 min, respectively (Fig. 4). Retention times and mass spectra for individual

standard molecules in solution need to be obtained in the same conditions. Degradation of the *N*-AHSL is evidenced by the reduction of the surface of the *N*-AHSL characteristic peak and concomitant increase in the surface of the *N*-AHS peak.

5. The identification of the degradation products is confirmed by mass spectrometry in comparison with those of the synthetic *N*-AHSL and *N*-AHS standards subjected to the same HPLC-MS/MS conditions. The specific fragments expected to appear in the mass spectra of *N*-AHSL, and its corresponding *N*-AHS should differ in size by one water molecule, e.g., 18 units (see example in Fig. 4).

3.5. Detection of N-AHSL Acylase Activities by HPLC (15)

1. The scheme used to detect the degradation products of *N*-AHSL following amidohydrolysis involves the chemical trapping of the free amine of the newly formed HSL (Fig. 5). As a consequence, it is best to use crude or purified protein extracts at this step (see Note 16).
2. Into a clean 2-mL microcentrifuge tube, add 50 μL of a 1 mM *N*-AHSL stock solution. Evaporate the ethyl acetate to dryness.
3. Add 500 μL of crude bacterial cell extract to the tube. Vortex for 1 min to dissolve the *N*-AHSL.
4. Incubate for up to 6 h at 25°C (see Note 8).
5. Add 25 μL of dansyl chloride solution to a final concentration of 185 mM.
6. Perform a control reaction with synthetic HSL in the same reaction conditions.
7. Incubate for 1 h at room temperature with frequent shaking (see Note 17).
8. Extract the reaction with one volume of dichloromethane. Transfer the upper phase to a clean tube and evaporate to dryness.

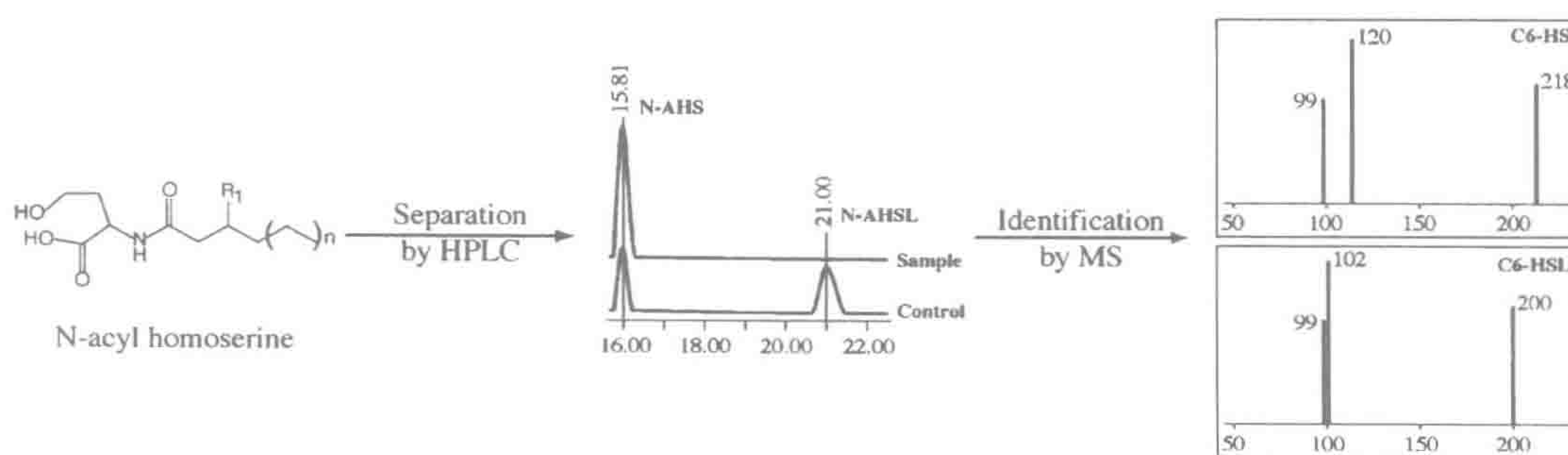


Fig. 4. Identification scheme for *N*-AHSL lactonase activities.

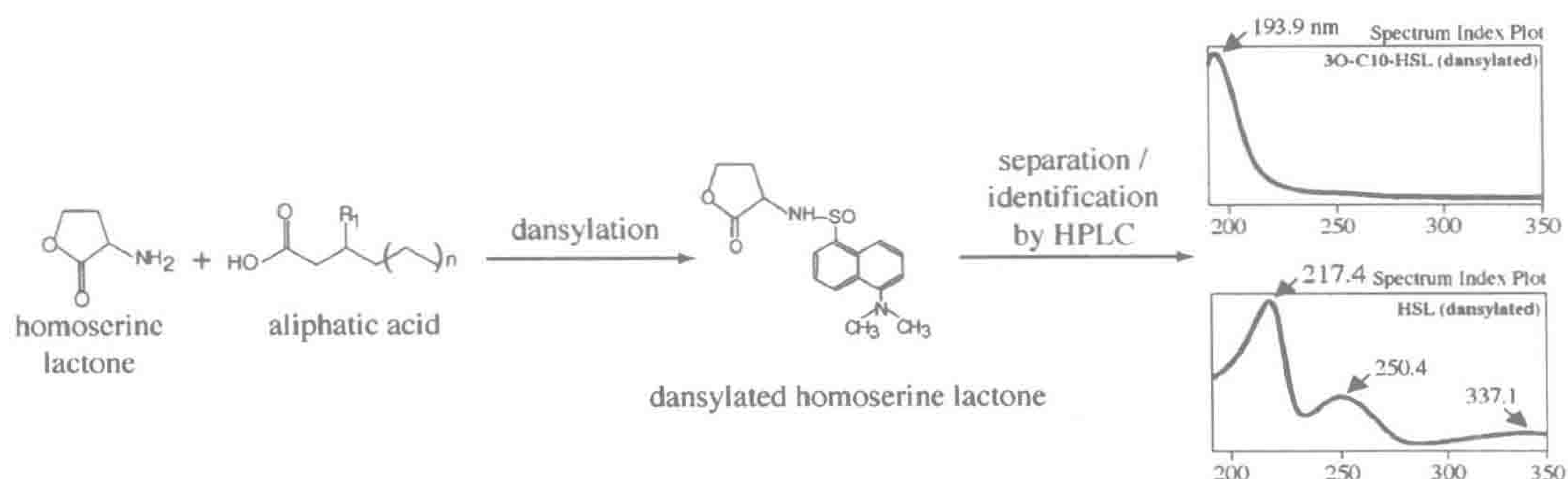


Fig. 5. Identification scheme for *N*-AHSL amidohydrolase activities.

9. Dissolve in 50 μL HCl 0.2 N to hydrolyse the excess of dansylchloride.
10. Extract with 50 μL of acetone.
11. Inject 10 μL of the reaction mixture in the Waters 625 HPLC system equipped with the C8 column. Detection of the dansyl moieties is performed with the Waters 996 PDA photodiode array detector.
12. Elution of the sample is performed in isocratic conditions with acetonitrile/water (35% acetonitrile in water) over a 30 min period at a flow rate of 2 mL/min. In these experimental conditions, dansylated-homoserine lactone harbors a retention time of 6.5 min. The confirmation of identification of the dansylated-homoserine lactone is obtained by plotting the spectrum index plot and comparing with the spectra obtained for the control reaction (Fig. 5).

4. Notes

1. Since the oxidoreduction of *N*-AHSLs generates alternate forms of the *N*-AHSL molecules, they have a lesser biotechnological potential. As a consequence of their retaining biological activity, the procedure described here will be inadequate to screen for these activities.
2. Several other sensors strains based on the same or different QS systems are available (18). The procedure presented here could easily be adapted for these sensors.
3. If one does not have a homemade container, a single-use container can be made out by taping all four sides of the TLC plate. Make sure that the plates are properly sealed to avoid leaks. This is easily obtained if the tape also covers part of the

reverse side of the TLC plate. This system does not allow the removal of bubble or the adjustment of the overlay thickness.

4. This screen can be used to screen virtually any type of microbial cell type/protein, from bacteria to fungi, from wild-type strains to clones overexpressing cloned *N*-AHSL degradation genes, from growing cells to purified proteins. Thus, it may be necessary to adjust the incubation time for each condition.
5. The procedure to follow to assay for "false positives" isolated in the microplate assay is essentially the same except that one just needs to run the original degradation reaction after extraction with 1 volume of ethyl acetate. Then, proceed directly to step 10.
6. It is recommended to evaporate the *N*-AHSLs under a flux of nitrogen to avoid chemical alteration.
7. The same procedure can be followed with CCE or purified *N*-AHSL lactonases. Incubation times and buffer conditions may have to be adapted to reflect these systems.
8. The incubation time should be optimized for each system by performing kinetic experiments.
9. The quantity of *N*-AHSL necessary for an easy detection by TLC using the different *N*-AHSL sensor systems will differ from one sensor to the other. Reference concentrations for each *N*-AHSL can be found for the *Agrobacterium* and *Chromobacterium* sensor system in refs. (17) and (13), respectively.
10. The TLC plate assay is easily adaptable to other sensor systems. To use it with the *C. violaceum* sensor CV026, proceed as noted above with the following modifications. The sensor culture is a 5 mL culture of CV026 grown overnight at 30°C. The overlay is composed of LB soft (7 g/L) agar (150 mL) to which the sensor culture is added.
11. The concentration of recircularized *N*-AHSL is difficult to estimate, since *N*-AHS might be further metabolized by some microorganisms. Thus, it might be necessary to spot different volumes of the acidified sample. No more than 5 μL of sample should be spotted to limit the diffusion of the product in the TLC.
12. We have noticed that some batch of TLC plates have a tendency to form bubbles at the interface between the TLC layer and overlay. It is important to remove those bubbles because they prevent the contact between the overlay and TLC and thus the transfer of the *N*-AHSL into the overlay and subsequently the induction of the sensor. These bubbles may be

removed with extreme care to avoid damaging the layer of the TLC using a small rounded spatula.

13. Square 245 mM culture dishes are the most convenient. Dishes can be reused.
14. The TLC approach also has the potential to identify the oxidoreductase and oxidase activities since the Rf and shape of the spot of the QS-active derivatives are different from the starting *N*-AHSL. In this case, spots with specific Rf and shapes different from the *N*-AHSL would be seen in both A and NA lanes. However, clones harboring such activities do not pass the microplate screening step because they generate QS-active derivatives.
15. Elution conditions may need to be adapted for the best separation of different *N*-AHSLs.
16. The same procedure can be followed with purified proteins but may require adaptation of buffer and incubation times.
17. The optimal temperature for the dansylation reaction is 37°C. However, incubating at this temperature favors the opening of the lactone ring. Thus, it is preferable to incubate at room temperature, although the reaction is less efficient.

References

1. Winans, S.C. and Bassler, B.L. (2002) Mob psychology. *J Bacteriol* 184, 873–883.
2. Reading, N.C. and Sperandio, V. (2006) Quorum sensing: the many languages of bacteria. *FEMS Microbiol Lett* 254, 1–11.
3. Hassett, D.J., Ma, J.F., Elkins, J.G., McDermott, T.R., Ochsner, U.A., West, S.E., et al. (1999) Quorum sensing in *Pseudomonas aeruginosa* controls expression of catalase and superoxide dismutase genes and mediates biofilm susceptibility to hydrogen peroxide. *Mol Microbiol* 34, 1082–1093.
4. Beck von Bodman, S. and Farrand, S.K. (1995) Capsular polysaccharide biosynthesis and pathogenicity in *Erwinia stewartii* require induction by an N-acylhomoserine lactone autoinducer. *J Bacteriol* 177, 5000–5008.
5. Rasmussen, T.B. and Givskov, M. (2006) Quorum sensing inhibitors: a bargain of effects. *Microbiology* 152, 895–904.
6. Uroz, S., Dessaux, Y., and Oger, P. (2009) Quorum sensing and quorum quenching: the yin and yang of bacterial communication. *ChemBioChem* 10, 205–216.
7. Dong, Y.H., Xu, J.L., Li, X.Z., and Zhang, L.H. (2000) AiiA, an enzyme that inactivates the acylhomoserine lactone quorum-sensing signal and attenuates the virulence of *Erwinia carotovora*. *Proc Natl Acad Sci USA* 97, 3526–3531.
8. Finch, R.G., Pritchard, D.I., Bycroft, B.W., Williams, P., and Stewart, G.S. (1998) Quorum sensing: a novel target for anti-infective therapy. *J Antimicrob Chemother* 42, 569–571.
9. Yang, F., Wang, L.H., Wang, J., Dong, Y.H., Hu, J.Y., and Zhang, L.H. (2005) Quorum quenching enzyme activity is widely conserved in the sera of mammalian species. *FEBS Lett* 579, 3713–3717.
10. Uroz, S., Oger, P., Chapelle, E., Adeline, M.-T., Faure, D., and Dessaux, Y. (2008) A *Rhodococcus qsd*A-encoded enzyme defines a novel class of large-spectrum quorum-quenching lactonases. *Appl Environ Microbiol* 74, 1357–1366.
11. Lin, Y.H., Xu, J.L., Hu, J., Wang, L.H., Ong, S.L., Leadbetter, J.R., et al. (2003) Acyl-homoserine lactone acylase from *Ralstonia* strain XJ12B represents a novel and potent class of quorum-quenching enzymes. *Mol Microbiol* 47, 849–860.
12. Luo, Z.Q., Su, S., and Farrand, S.K. (2003) *In situ* activation of the quorum-sensing

transcription factor TraR by cognate and noncognate acyl-homoserine lactone ligands: Kinetics and consequences. *J Bacteriol* 185, 5665–5672.

13. McClean, K.H., Winson, M.K., Fish, L., Taylor, A., Chhabra, S.R., Camara, M., et al. (1997) Quorum sensing and *Chromobacterium violaceum*: exploitation of violacein production and inhibition for the detection of *N*-acylhomoserine lactones. *Microbiology* 143, 3703–3711.
14. Reimmann, C., Ginet, N., Michel, L., Keel, C., Michaux, P., Krishnapillai, V., et al. (2002) Genetically programmed autoinducer destruction reduces virulence gene expression and swarming motility in *Pseudomonas aeruginosa* PAO1. *Microbiology* 148, 923–932.
15. Uroz, S., Chhabra, S.R., Camara, M., Williams, P., Oger, P., and Dessaux, Y. (2005) N-Acylhomoserine lactone quorum-sensing molecules are modified and degraded by *Rhodococcus erythropolis* W2 by both amidolytic and novel oxidoreductase activities. *Microbiology* 151, 3313–3322.
16. Yates, E.A., Philipp, B., Buckley, C., Atkinson, S., Chhabra, S.R., Sockett, R.E., et al. (2002) N-acylhomoserine lactones undergo lactonolysis in a pH-, temperature-, and acyl chain length-dependent manner during growth of *Yersinia pseudotuberculosis* and *Pseudomonas aeruginosa*. *Infect Immun* 70, 5635–5646.
17. Shaw, P.D., Ping, G., Daly, S.L., Cha, C., Cronan, J.E., Jr., Rinehart, K.L., et al. (1997) Detecting and characterizing N-acyl-homoserine lactone signal molecules by thin-layer chromatography. *Proc Natl Acad Sci U S A* 94, 6036–6041.
18. Winson, M.K., Swift, S., Fish, L., Throup, J.P., Jorgensen, F., Chhabra, S.R., et al. (1998) Construction and analysis of *lux*CD-ABE-based plasmid sensors for investigating N-acyl homoserine lactone-mediated quorum sensing. *FEMS Microbiol Lett* 163, 185–192.

Chapter 15

Identification of Molecular Markers to Follow Up the Bioremediation of Sites Contaminated with Chlorinated Compounds

Massimo Marzorati, Annalisa Balloi, Francesca De Ferra, and Daniele Daffonchio

Abstract

The use of microorganisms to clean up xenobiotics from polluted ecosystems (soil and water) represents an ecosustainable and powerful alternative to traditional remediation processes. Recent developments in molecular-biology-based techniques have led to rapid and sensitive strategies for monitoring and identifying bacteria and catabolic genes involved in the degradation of xenobiotics. This chapter provides a description of recently developed molecular-biology-based techniques, such as PCR with degenerate primers set, real-time quantitative PCR (qPCR), reverse transcription PCR (RT-PCR), southern blot hybridization, and long-range PCR, used to give a picture of the catabolically relevant microorganisms and of the functional genes present in a polluted system. By using a case study of a groundwater aquifer contaminated with 1,2-dichloroethane (1,2-DCA), we describe the identification of microorganisms potentially involved in the 1,2-DCA dehalorespiration (*Dehalobacter* sp. and *Desulfitobacterium* sp.) and a complete new gene cluster encoding for a 1,2-DCA reductive dehalogenase. The application of these techniques to bioremediation can improve our understanding of the inner mechanisms to evaluate the feasibility of a given treatment and provide us with a method to follow up bacteria and catabolic genes involved in the degradation of contaminants during the activities in situ.

Key words: Reductive dehalogenase, Functional genes, *Dehalobacter* sp., *Desulfitobacterium* sp., 1,2-DCA, Dehalorespiration, Bioremediation

1. Introduction

The role of microorganisms in the remediation of polluted environments has been clearly recognized and their use to clean up xenobiotics represents a potential solution to such a problem. It has been demonstrated that a wide taxonomic and functional

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668, DOI 10.1007/978-1-60761-823-2_15,

bacterial diversity can be strictly correlated to the reclamation of polluted sites (1). Hence, bacteria monitoring becomes a key step in predicting and following the effectiveness of bioremediation processes – *in situ* as well as *ex situ* – during both natural and assisted attenuation procedures (2). Besides, recent developments in molecular biology techniques have led to rapid and sensitive strategies for monitoring and identification of bacteria involved in the degradation of contaminant compounds (3). Using these techniques, it is possible to investigate the biological components of a given polluted environment and to understand if the native microbiota is proper to support a remediation treatment. The combination of molecular-ecology-based techniques to investigate the presence of known microbial degraders (e.g., by monitoring 16S rRNA genes) and of genes codifying for the catabolic enzymes represent a very effective approach both to establish the feasibility of a treatment and to follow up an ongoing bioremediation treatment (4).

The concept of functional diversity and functional redundancy is a basic component in every remediation strategy (5). Among the several contaminants that are normally retrieved in the groundwater all over the world, chlorinated compounds (CCs) represent a major pollution problem being hazardous to humans and recalcitrant to degradation. CCs can be used by microorganisms as electron donors and carbon sources under anaerobic or aerobic conditions. In this chapter, we use 1,2-dichloroethane (1,2-DCA) as a reference compound to investigate the metabolic process of halorespiration, in which the CCs can serve as electron acceptors and are reductively dechlorinated. This dechlorination process is coupled with anoxic decomposition of simple organic substrate (H_2 is the final electron donor) and ethene production (6, 7).

Here, we present and discuss the procedures to identify molecular markers, in terms of 16S rRNA and reductive dehalogenase (RD) gene sequence diversity and abundance, to follow up a bioremediation treatment for 1,2-DCA dechlorination in a polluted groundwater (8). Based on our experience on 1,2 DCA, we aim to propose the efficacy and the sensitivity of the described experimental approach to monitor and manage bioremediation processes of any contaminated site.

2. Materials

2.1. Sampling

1. Pump system to collect water samples directly from the piezometers installed in the contaminated site.
2. Glass bottle to store water samples.
3. Multiparametric probe to measure depth, O_2, pH, and redox potential.

2.2. Microcosms Preparation

1. 50 mL penicillin flasks.
2. Solution of 1,2-DCA.
3. 1 mM cysteine solution.
4. 50 mg/L vitamin B12.
5. 0.5 mM Hepes/NaOH, pH 7 [4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid solution].
6. Yeast extract 0.05% (w/v).
7. 1:200 dilution of a trace elements stock solution containing 12.8 g/L nitrilotriacetic acid, 1.35 g/L $FeCl_3 \cdot 6H_2O$, 0.1 g/L $MnCl_2 \cdot 4H_2O$, 0.024 g/L $CoCl_2 \cdot 6H_2O$, 0.1 g/L $CaCl_2 \cdot 2H_2O$, 0.1 g/L $ZnCl_2$, 0.025 g/L $CuCl_2 \cdot 2H_2O$, 0.01 g/L H_3BO_3, 0.024 g/L $Na_2MoO_4 \cdot 2H_2O$, 1 g/L NaCl, 0.12 g/L $NiCl_2 \cdot 6H_2O$, and 0.026 g/L $Na_2SeO_3 \cdot 5H_2O$.
8. Salt solution containing final concentrations of 43 mg/L NH_4Cl, 0.5 g/L KH_2PO_4, 0.2 g/L $MgCl_2 \cdot 6H_2O$, and 0.01 g/L $CaCl_2 \cdot 2H_2O$.
9. Electron donors/carbon sources: Na-lactate, Na-acetate, Na-formate, or cheese whey.

2.3. Measurement of CCs Degradation and Chloride Concentration

1. Gas chromatograph with a flame ionization detector (GC-FID) (Varian) and a capillary column (Varian FactorFourTM Low Bleed VF 624 ms, 30 m × 0.25 mm).
2. 7694 Agilent gas chromatograph equipped with FID on a DB624 column (J&W Scientific, Folsom, CA) to measure the concentrations of 1,2-DCA, vinyl chloride, and other possible degradation products.
3. PC spectrophotometer (Lovibond, Dortmund, Germany) for chloride determination by utilization of method 180 chloride concentration (Tintometer GmbH, Dortmund, Germany).

2.4. Nucleic Acid Extraction and cDNA Synthesis

2.4.1. DNA Extraction

1. Sterivex filters (Millipore, Milan, Italy).
2. Lysis solution containing 1 mg/mL lysozyme, 1% (w/v) sodium dodecyl sulfate, and 0.5 mg/mL proteinase K.
3. Phenol–chloroform–isoamylalcohol (25:24:1).
4. Chloroform–isoamylalcohol (24:1).
5. TE: 10 mM Tris–HCl, 0.1 mM EDTA, pH 8.0.

2.4.2. RNA Extraction and cDNA Synthesis

1. RNA extraction using the NucleoSpin RNA II kit (Macherey-Nagel, Düren, Germany).
2. DNAse enzyme (Promega, Milan, Italy).
3. RevertAid™ H Minus M-muLV Reverse Transcriptase Kit (Promega).

4. Sequence-specific primer DHLR2 (5′GTAAACTTTCCCC GTCGC) obtained after the identification of the RD gene (see Subheading 3.6).

2.5. PCR and Quantitative PCR

All the PCR reactions were conducted in a Thermal cycler MyiQ (Bio-Rad, Milan, Italy); the quantitative PCR (qPCR) was performed with a GeneAmp 5700 RT PCR instrument (Applied Biosystems, Milan, Italy).

2.5.1. 16S rRNA Gene PCR

1. 16S rRNA gene has been amplified from the groundwater metagenome using bacterial universal primers 27f (5′AGAG TTTGATCCTGGCTCAG) and 1494r (5′CTACGGCTACC TTGTTACGA).
2. Platinum® Taq DNA Polymerase High Fidelity PCR kit (Invitrogen, Milan, Italy).

2.5.2. Identification of a RD Gene

1. Degenerated primers ceRD2Sf (5′GCAGCACGCCTTTTT GGIGCIKMIWSIGTIGG), ceRD2Lf (5′GCAGCACGCCT TTTTGGIGCIKMIYTNGTIGG), and RD7r (5′AANGGRC AIACIGCIWCRCA) (9).
2. Platinum® Taq DNA Polymerase High Fidelity PCR kit (Invitrogen, Milan, Italy).
3. QIAquick PCR Purification Kit (Qiagen, Milan, Italy).

2.5.3. Inverse and Direct PCR

1. *Eco*RI restriction enzyme (Promega, Milan, Italy) and relative 10× buffer H (Promega).
2. Sterile Tris–EDTA (pH 8) buffer: 10 mM Tris–HCl, 0.1 mM EDTA, pH 8.0.
3. Phenol–chloroform–isoamylalcohol (25:24:1).
4. Na-acetate, 70% and 100% ethanol.
5. Sterile MilliQ water.
6. T4 DNA ligase (Promega) and 10× ligation buffer (Promega).
7. RevertAid™ M-MuLV Reverse Transcriptase (Fermentas, Milan, Italy).
8. Primers in Table 1.

2.5.4. RD-Specific Quantitative PCR

1. GeneAmp® 5700 RT PCR instrument (Applied Biosystems, Milan, Italy).
2. Primers DH3F (5′ATTGGGAGAAGCATGCAGGT3′) and DH3R (5′GACCACCGTTATAGGCCCAGA3′) specific for *dcaA* gene.
3. DyNAmoTM HS SYBR® Green qPCR (Celbio, Milan, Italy).

2.5.5. RD Expression

1. Primers DHLF1 (5′GGACCTCGTTGGACTCC3′), DHLF2 (5′GTTAAAAAGGCAGCCTGTT3′), DHLR1 (5′GGCAAA TCCCATGGCATTA 3′), and DHLR2 (5′GTAAACTTTCC CCGTCGC3′) specific for the *dcaA* gene, and DcaB rev (5′TGGTATTCACGCTCCGA3′) specific for the *dcaB* gene.
2. RevertAid™ M-MuLV Reverse Transcriptase (Fermentas, Milan, Italy).

2.5.6. RD Clone Library

1. Primers PceAFor1 (5′ACGTGCAATTATTATTAAGG3′) and DcaB rev (Table 1).
2. RevertAid™ M-MuLV Reverse Transcriptase (Fermentas, Milan, Italy).
3. pGEM cloning kit (pGEM-T Easy Vector Systems, Promega, Milan, Italy).

2.6. 16S rRNA Gene Libraries

1. pGEM cloning kit (pGEM-T Easy Vector Systems, Promega, Milan, Italy).
2. Primers T7 (3′CTAATACGACTCACTATAGGG5′) and SP6 (3′ATTTAGGTGACACTATAGAATA5′).
3. QIAquick PCR Purification Kit (Qiagen, Milan, Italy).
4. Primer 27F (5′AGAGTTTGATCCTGGCTCAG).
5. ABI Prism BigDye terminator cycle sequencing kit (Applied Biosystems, Milan, Italy).
6. ABI 310 automated sequencer (Applied Biosystems).

2.7. Southern Blot

1. Two probes specific for *dcaA* and *dcaB* of the RD cluster and labeled with digoxigenin (DIG) by random priming.
2. Probe A has been produced by PCR using primers Dca1F and DHL-REV (Table 1) amplifying a variable region of the gene A; while primers Dca1BF and Dca1BR (Table 1) have been used to amplify a probe in a conserved region of the gene B (probe B).
3. Labeling, prehybridization, hybridization, and detection were performed with the DIG DNA labeling and detection kit (Boehringer Mannheim, Milan, Italy) according to the manufacturer's instructions.
4. Enzyme *Eco*RI and relative buffer.
5. Denaturation solution: 0.5 M NaOH, 1,5 M NaCl.
6. Neutralization solution: 0.5 M Tris–HCl (pH 7.2), 1.5 M NaCl, 1 mM NaCl.
7. 20× SSC: 3 M NaCl, 0.3 M Na Citrate, adjust pH to 7.4 with 1 M HCl.
8. Standard prehybridization buffer: 5× SSC, 2.0% blocking reagent (w/v).

Table 1
Summary of all the primers used for the inverse PCR, direct PCR, and sequencing of the environmental RD and RD-DCA1 of *D. dichloroeliminans* strain DCA1 responsible for 1,2-DCA reductive dehalogenation (8)

Primer	DNA used for PCR	5′–3′ sequence	Orientation	nt position[a]
DHL-FOR	Env RD	AATTCGGGGTACGCGAGT	For	2771
DHL-REV	Env RD	CAGGCTCATTAGCTATTTCA	Rev	2097
DHL-for a	Env RD–DCA1	TCGGAGCGTGAATACCA	For	3429–3306
DcaB-Rev	Env RD–DCA1	TGGTATTCACGCTCCGA	Rev	3429–3306
DHL-RevA	Env RD	TACTTTGCATCCACCTTG	Rev	1696
PceT-rev2	Env RD	GATTAACTTGCCAAATTGAT	Rev	6021
Dehalo-PCE	Env RD	ATAATGACTCAACTTTCGAA	For	0
DHL-ups-Mid	Env RD	TGTGTAGGAGTTACGACA	For	663
PceT-F	Env RD	GTATGAATTTGATGAAGAAG	For	5645
DHL-orf-Rev	Env RD	AGGAGAGAACCTTAATCG	Rev	7460
DHL-fin-Rev	Env RD	GGATTTGTTCCTCATCCT	Rev	6824
PceAFor1	Env RD–DCA1	ACGTGCAATTATTATTAAGG	For	1485–1379
DcaC-F2	Env RD–DCA1	ATTCTTCTGTCCGGTAGGAT	For	4808–4702
Pce-C-rev2	Env RD–DCA1	TACCTCCTCATTTCGCC	Rev	4852–4746
DCA1ups	DCA1	ATGCAAAGCTAGGTGCTG	For	0
tpnDCA1rev	DCA1	TACAGTGTCCCCATCCG	Rev	86
DCA1dwnR	DCA1	AGAGACTGGTGTAGGTTG	Rev	498
RDdca1R1	DCA1	TCCTCCTGTTGATTTCGC	Rev	1415
RDdca1F1	DCA1	AATACCTTGTTGGATGACG	For	1843
RDdca1R2	DCA1	TTGGTTTGAAACCCGCATA	Rev	2311
dcaTfwDCA1	DCA1	TCTTGAAAAGCGAATTAACG	For	5783
DCA1dcaTr	DCA1	GATTAACTTGCCAAATTGATT	Rev	5907

[a]The nucleotide position refers to the position of the primer in the complete sequence of the RD samples they have been used on. Where there are two numbers, the first one refers to Env RD and the second one to RD-DCA1

9. Blocking reagent (2%): 0.1% *N*-lauroylsarcosine, 0.02% SDS (sodium dodecyl sulfate), 50% formamide.
10. Standard hybridization buffer: DIG-labeled probes diluted in standard prehybridization buffer.
11. Washing solution I: 2× SSC, 0.1% SDS.
12. Washing solution II: 0.1× SSC, 0.1% SDS.

13. Buffer 1: 4× (stock): 0.6 M NaCl, 0.4 M maleic acid, adjust pH to 7.5 NaOH.
14. Buffer 2: 1× buffer 1, 1% (w/v) blocking reagent.
15. Buffer 3: 0.1 M Tris–HCl (pH 9.5), 0.1 M NaCl, 50 mM $MgCl_2$.
16. Washing buffer: 1× buffer 1, 0.3% Tween 20.
17. CSPD® chemiluminescent substrate, ready to use (Roche Diagnostics SpA Monza, Milan, Italy).
18. Whatman 3 mM paper.
19. Boehringer Mannheim's Nylon membrane, positively charged.

3. Methods

3.1. Sampling

1. In the contaminated site, take measurements with the multiparametric probe.
2. Apply the sucking device of the pump below the water level.
3. Collect the water samples in bottles with no headspace in order to minimize oxygen contamination.
4. Either process the samples immediately or store them at room temperature in the dark for up to a few days before setting up the microcosms.

3.2. Microcosms Preparation

Different possible conditions are tested in laboratory with microcosms to identify the best electron donor in supporting the 1,2-DCA reductive dehalogenation. From these experiments, lactate resulted to be the best electron donor supporting the reductive dehalogenation of 900 ppm of 1,2-DCA in 15 days.

1. Each microcosm is set up in a sterile penicillin flask.
2. In all the microcosms, 10 or 50 ppm of 1,2-DCA are added after purging the water with sterile oxygen-free nitrogen.
3. All the microcosms are amended with cysteine (to help maintaining the anaerobic condition), vitamin B12 (a basic cofactor for some dehalogenating microorganisms), Hepes/NaOH (to keep the pH in a range of 6.8–8.2), yeast extract, a trace elements stock solution and a supplementary salt solution.
4. Alternatively, the microcosms can be provided with Na-lactate or Na-acetate (5 mM both), Na-formate (40 mM) or cheese whey (0.2% w/v) as carbon source/electron donors.
5. Microcosms amended with cheese whey could not receive the supplementary salt solution.
6. Control microcosms are prepared by incubating parallel vials containing all amendments with filter-sterilized groundwater

samples or additional 0.1% (w/v) benzalconium chloride as bactericide.

7. All microcosms must be immediately sealed after addition of 1,2-DCA with Teflon-faced septa and aluminum crimp seals, and incubated in the dark at 23°C.

3.3. Measurement of CCs Degradation

In order to follow up the 1,2-DCA reductive dechlorination, ethene, vinyl chloride, and other possible 1,2-DCA degradation products are analyzed by headspace gas chromatography.

1. Headspace sample of 1 mL is injected in a gas chromatograph with a flame ionization detector (GC-FID) and a capillary column Varian FactorFourTM Low Bleed VF 624 ms, 30 m × 0.25 mm.
2. The temperature of the injector is 200°C with a split ratio of 5.
3. The temperature program starts at 30°C for 2 min, followed by an increase to 40°C at 1°C/min and afterwards to 105°C at 5°C/min.
4. Helium is used as the carrier gas at a flow of 3 mL/min.
5. The temperature of the detector is 250°C.
6. 1,2-DCA limit of detection is about 1–2 μg/L.

The chloride concentration is measured on a PC spectrophotometer by utilization of method 180 chloride concentration. Chloride range of detection is between 5 and 60 mg/L.

3.4. Nucleic Acids Extraction and cDNA Synthesis

3.4.1. DNA Extraction

1. Groundwater (30 mL) and microcosm water (1.5 mL) samples are filtered using Sterivex filters to concentrate bacterial cells for the DNA extraction.
2. DNA is extracted from the filtered bacterial cells by incubating the filter with 2 mL of a lysis solution to facilitate the release of the nucleic acids in the solution.
3. The DNA is purified as previously described by Murray et al. (10):
   (a) Transfer the supernatant in two clean 2 mL eppendorf tube.
   (b) Add 1 volume of phenol–chloroform–isoamylalcohol (25:24:1).
   (c) Add 1 volume of chloroform–isoamylalcohol (24:1).
   (d) Add 1 volume of isopropanol 100% (1 h at room temperature) to precipitate the DNA.
   (e) Centrifuge 25 min at 15°C.
   (f) Resuspend the DNA pellets in 100 μl of TE (pH 8).

3.4.2. RNA Extraction and cDNA Synthesis

1. RNA extraction is performed on a 8.5 mL sample from a dechlorinating microcosms by using the NucleoSpin RNA II kit according to the manufacturer's instructions. A higher volume is used, compared to the DNA extraction, to increase the final amount of purified RNA.
2. Total extracted RNA is treated with 10 U of DNAse at 37°C for 15 min to remove any possible trace of contaminant DNA.
3. cDNA is synthesized using the RevertAid™ H Minus M-muLV Reverse Transcriptase Kit as follows:
   3.1. The reverse transcription reaction is performed in a mixture containing 0.3 μg of total RNA, 20 pmol of the primer DHLR2 (a reverse primer is always needed for annealing to mRNA), and deionized water (nuclease free) to the final volume of 11 μl.
   3.2. After incubation at 70°C for 5 min and chilling on ice, 5× reaction buffer, 1 mM (final concentration of each dNTP) dNTP mix, and deionized water (nuclease free) to the final volume of 19 μl are added to the mix.
   3.3. After 5 min at 37°C, the reverse transcription is conducted adding 200 units of M-MuLV reverse transcriptase and incubating the mix at 42°C for 60 min and 70°C for 10 min to stop the reaction.
4. Extracted RNA is ready to be used to investigate the expression of the RD gene cluster (see Subheading 3.8).

3.5. 16S rRNA PCR and Gene Library

1. 16S rRNA genes of the microbial community are amplified from the groundwater metagenome using the bacterial universal primers 27f and 1494r.
2. Reaction: 1× PCR buffer, 1.5 mM $MgCl_2$, 0.12 mM dNTPs, 0.3 μM of each primer, 1 U of Taq polymerase in a final volume of 50 μl.
3. Thermal protocol: initial denaturation at 94°C for 5 min, followed by 5 cycles consisting of denaturation at 94°C for 1 min, annealing at 50°C for 1 min, and extension at 72°C for 2 min, and by 30 cycles consisting of denaturation at 94°C for 1 min, annealing at 55°C for 1 min, and extension at 72°C for 2 min. A final extension at 72°C for 10 min is added.
4. Use 60 ng of PCR product (insert:vector molar ratio of 3:1) for cloning reactions using the pGEM cloning kit following the recommendations of the manufacturer.
5. A direct PCR assay is performed on white colonies (following X-gal selection) to amplify the cloned inserts using primers T7 and SP6.

6. Amplification reactions are performed in a 25 μl total volume containing: 10× reaction buffer, 0.12 mM of each dNTP, 0.7 U of Taq DNA polymerase, and 0.3 μM of each primer.
7. Reactions are performed with the following protocol: an initial melting at 94°C for 4 min, followed by 30 cycles at 94°C for 30 s, 50°C for 1 min, 72°C for 1.5 min, and a final extension step at 72°C for 5 min.
8. PCR products are then purified by the QIAquick PCR Purification Kit according to the manufacturer's instructions, in preparation of the sequencing reaction.
9. Clones are sequenced with primers 27f on an ABI 310 automated sequencer.
10. Rarefaction curves are built using the PAST program to observe the percentage of coverage of the microbial total diversity (11).
11. To determine the operational taxonomic units (OTUs), the 99% identity criterion for the full-length 16S rRNA sequence can be chosen to not underestimate the true diversity of the ecosystem (12).

3.6. RD Identification and "Inverse and Direct" PCR Approach

The entire procedure here presented led to the identification of a new RD specific for 1,2-DCA with a common gene cluster organization as reported by Maillard et al. (13) and Marzorati et al. (8): *tpnA* that codifies for a transposase; *dcaA* that encodes the RD; *dcaB* for a membrane protein; *dcaC* for a protein involved in the regulation of the membrane–enzyme interaction; *dcaT* for a trigger factor that probably activates the RD in presence of the contaminant.

1. To identify potential catabolic genes involved in reductive dehalogenation, PCR amplification with degenerated primers (ceRD2Sf, ceRD2Lf, RD7r) is performed on DNA extracted from groundwater microcosms according to Regeard et al. (9).
2. Amplification reactions are performed in a volume of 25 μl containing 2.5 μl of 10× reaction buffer, 2 μl of 2.5 mM of each dNTP, 2.5 units of Taq DNA polymerase, and 1 μl of each degenerate primer (stock solution at 20 μM).
3. Thermal protocol: 3 min initial denaturation at 94°C, 36 cycles of 1 min denaturation at 94°C, 1 min of primer annealing at 47°C, 1 min of elongation at 72°C. A final extension step of 10 min at 72°C is included.
4. The new sequences are aligned with those already described in the NCBI international database to identify the presence of conserved or divergent regions in the new putative RD genes.
5. The flanking regions of the newly identified fragment of the catalytic subunit are identified by inverse PCR. For this purpose,

250 ng of metagenomic DNA are digested with *Eco*RI restriction enzyme for 15 h at 37°C in a 20 µl reaction containing 2 µl of 10× buffer H and 15 U of the restriction enzyme.

6. After the digestion, 30 µl of sterile Tris–EDTA (pH 8) buffer is added to the digestion mix, the enzyme inactivated with 1 volume of phenol–chloroform–isoamylalcohol (25:24:1) and the DNA precipitated with 5 µl of Na-acetate and 2 volumes of 100% ethanol.
7. After centrifugation at 13,000×*g* for 10 min, the DNA pellet is rinsed with 250 µl of 70% ethanol and resuspended with 40 µl of sterile MilliQ water.
8. Five units of T4 DNA ligase, 5 µl of the 10× ligation buffer (Promega), and MilliQ water for a final volume of 50 µl are then added to the digested genomic DNA.
9. The ligation mixture is incubated for 1 h at 22°C – to allow the fragments generated at step 6 to self-ligate (Fig. 1) – and then 10 min at 65°C to inactivate the enzyme.
10. Three microliters of the ligation mix is then used as a template for the following PCR reaction: 1× PCR buffer, 1.5 mM $MgCl_2$, 0.3 mM dNTPs, 0.5 µM of each primer (DHL-FOR/DHL-REV), 1 U of Taq polymerase in a final volume of 50 µl.
11. Thermal protocol: initial denaturation at 94°C for 4 min, followed by 30 cycles consisting of denaturation at 94°C for 45 s, annealing at 58°C for 1 min, and extension at 72°C for 2 min and 15 s. A final extension at 72°C for 7 min is added.
12. Subsequently, direct PCR experiments to cover the entire gene cluster are conducted (based on the fact the genes codifying for the RD structural subunits are highly conserved) with the following reaction: 1× PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.3 µM of each primer, 1 U of Taq polymerase in a final volume of 50 µl. The following couples of primers are applied: DHL-for a/PceTrev2; Dehalo-PCE/DHL-RevA; PceT-F/DHL-orf-Rev (Table 1).
13. Thermal protocol: initial denaturation at 94°C for 4 min, followed by 30 cycles consisting of denaturation at 94°C for 45 s, annealing at 52°C for 1 min, and extension at 72°C for about 1 min every 1,000 bp of expected product. A final extension at 72°C for 7 min is added.
14. Finally, to verify that indeed all the identified genes are part of a single gene cluster, a long-range PCR with primers annealing within *tpnA* and downstream *dcaT* should be performed to obtain a single PCR product of about 5,000 bp.

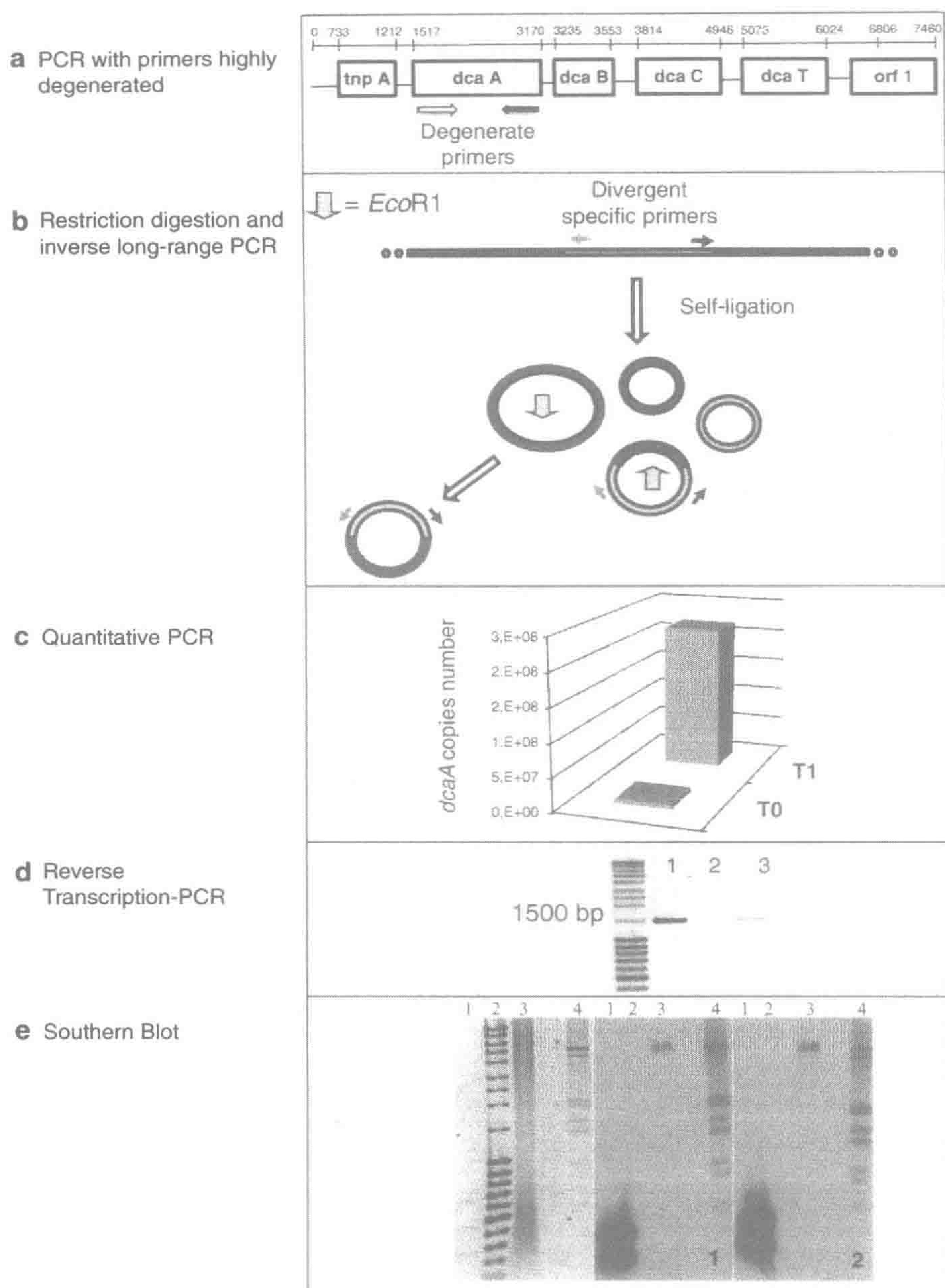


Fig. 1. Scheme of the procedure applied to identify a new RD cluster specifically adapted to 1,2-DCA reductive dechlorination. The new RD gene cluster has been identified by a step-by-step approach that includes (**a**) PCR with primers highly degenerated, (**b**) inverse PCR, direct and long-range PCR. The actual involvement of the new RD was evaluated by (**c**) quantitative PCR (the *dcaA* gene copies increased following the biostimulation treatment), (**d**) RT-PCR (the *dcaA* and *dcaB* genes were cotranscribed during the biostimulation treatment), and (**e**) Southern blot (the same RD cluster was identified in *Desulfitobacterium dichloroeliminans* strain DCA1).

3.7. Quantitative PCR to Evaluate the dcaA Gene Copy Number after a Simulated Biostimulation Treatment

1. To enumerate the copies of *dcaA* gene in the extracted environmental DNA, a quantitative PCR assay is conducted using the primers DH3F and DH3R specific for the *dcaA* gene.
2. Metagenomic DNA (200 pg) are used in 50 µl reaction with DyNAmoTM HS SYBR® Green qPCR according to manufacturer instructions.
3. Thermal protocol: initial hot start at 50°C for 2 min and denaturation at 94°C for 15 min were followed by 35 cycles consisting of denaturation at 94°C for 10 s, annealing at 55°C for 30 s, and extension at 72°C for 30 s. A final step for melting curve analysis from 72 to 95°C, measuring fluorescence every 0.5°C, is added.
4. Reference curves are run in every experiment by utilization of known amounts of clone RD-54 (14).

3.8. RDs Gene Expression

1. Starting from the cDNA synthesized in Subheading 3.4, the expression of the newly identified RD is evaluated using primers DHLF1, DHLF2, DHLR1, and DHLR2 specific for the *dcaA* gene, and DcaB rev specific for the *dcaB* gene.
2. Reaction: 1× PCR buffer (Amersham), 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.6 µM of each primer, 1 U of Taq polymerase in a final volume of 25 µl.
3. Thermal protocol: initial denaturation at 94°C for 3 min, followed by 31 cycles consisting of denaturation at 94°C for 30 s, annealing at 54°C for 1 min, and extension at 72°C for 1 min and 15 s. A final extension at 72°C for 7 min is added.

3.9. Southern Blot: Identification of the RD Gene Cluster in Desulfitobacterium dichloroeliminans Strain DCA1

The number of dehalogenase in the genomic DNA of the strain DCA1 has been investigated by Southern hybridization experiments with two probes specific for gene A and gene B of the RD cluster and labeled with digoxigenin (DIG) by random priming.

1. Labeling, prehybridization, hybridization, and detection are performed with the DIG DNA labeling and detection kit according to the manufacturer's instructions.
2. Depurinate DNA to facilitate transfer of large DNA.
3. Genomic DNA is digested with *Eco*RI (choose an enzyme that possibly does not cut within the RD gene cluster).
4. Load the digested DNA on an agarose gel with appropriate DNA size markers and stain with ethidium bromide.
5. The gel is then washed twice both in the denaturation and neutralization solutions.
6. Equilibrate the gel for 15 min in 20× SSC.
7. The DNA is then transferred overnight to a nylon membrane using the sterile 20× SSC as transfer buffer and Whatman 3 mM paper.

8. The nylon membrane is washed in 2× SSC, dried and then the DNA is UV-fixed on the membrane by means of a transilluminator for 5 min.
9. Incubate in a sealed plastic bag the nylon membrane with the prehybridization solution at 42°C; the hybridization is conducted overnight at 42°C in the presence of 50% (v/v) formamide; two 15-min stringent washes are performed at 65°C to increase the hybridization specificity.
10. After the hybridization step, the membrane is washed with the following solutions.
    (a) Washing solution I at room temperature for 15 min.
    (b) Washing solution II at 50–70°C.
    (c) Washing buffer at room temperature for 5 min.
11. The membrane is then incubated with the antibody solution for 30 min at room temperature.
12. The membrane is washed in the following solutions.
    (a) Washing buffer at room temperature for 5 min.
    (b) Washing buffer 3 at room temperature for 5 min.
13. The chemiluminescent detection is performed incubating the membrane in buffer 3 with CSPD chemiluminescent substrate at 37°C for 15 min.
14. The membrane is then exposed to X-ray film for 15–25 min at room temperature.

3.10. RDs Gene Library to Assess the Functional Redundancy

1. Primers PceAFor1 and DcaB rev are used to amplify a region including the entire *dcaA* gene and 194 bp of *dcaB* gene on *D. dichloroeliminans* strain DCA1 genomic DNA and on the environmental metagenome.
2. Reaction: 1× PCR buffer (Amersham), 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.6 μM of each primer, 1 U of Taq polymerase in a final volume of 25 μl.
3. Thermal protocol: initial denaturation at 94°C for 3 min was followed by 31 cycles consisting of denaturation at 94°C for 30 s, annealing at 54°C for 1 min, and extension at 72°C for 1 min and 15 s. A final extension at 72°C for 7 min is added.
4. The procedure to build the library is the same reported in Subheading 3.5, step 4.

4. Conclusions

1. As previously stated, the flow of the experiments aimed at the identification of possible molecular markers to investigate the

feasibility of a bioremediation treatment and, in a second phase to follow up the treatment in situ. First, the metabolic capabilities of the microbial community resident in the contaminated aquifer were evaluated in microcosms studies. The structure and composition of said microbial community were evaluated by means of 16S rRNA gene libraries prior and after a simulated biostimulation treatment.

2. The presence of genes correlated with the reductive dehalogenation of 1,2-DCA was investigated with highly degenerated primers. Once a putative fragment was identified, the flanking regions were obtained by means of a combination of inverse and direct PCR. This led to the identification of a new gene cluster. Evidences of the role of the new identified cluster in the 1,2-DCA degradation were obtained assessing (a) the increased gene copies number after the biostimulation treatment; (b) the expression of the gene coding for the catalytic subunit during the biostimulation treatment; and (c) the presence of the sole gene cluster in *D. dichloroeliminans* strain DCA1, a microorganism that can reductive dechlorinate 1,2-DCA into ethene with no more toxic chlorinated intermediates (8). The entire procedure is summarized in Fig. 1.
3. The capacity of sharing metabolic functionality between different members of a microbial community as well the presence of genetic redundancy is a fundamental feature for the efficiency of the community itself. This characteristic was investigated with RD gene clone libraries using primers specific for conserved regions of the RD gene cluster that could amplify the variable region (*dcaA*) that codes for the catalytic subunit.

Acknowledgements

The research was supported by a grant from Eni S.P.A. as part of the project "Caratterizzazione microbiologica e molecolare di siti contaminati da idrocarburi clorurati" (Contratto no. 4800040841).

References

1. Smidt, H. and de Vos, W.M. (2004) Anaerobic microbial dehalogenation. *Annu Rev Microbiol* 58, 43–73.
2. Van Raemdonck, H., Maes, A., Ossieur, W., Verthe, K., Vercauteren, T., Verstraete, W., et al. (2006) Real time PCR quantification in groundwater of the dehalorespiring *Desulfitobacterium dichloroeliminans* strain DCA1. *J Microbiol Meth* 67, 294–303.
3. Widada, J., Nojiri, H., and Omori, T. (2002) Recent developments in molecular techniques for identification and monitoring of xenobiotic-degrading bacteria and their catabolic genes in bioremediation. *Appl Microbiol Biotechnol* 60, 45–59.
4. Power, M., van der Meer, J.R., Tchelet, R., Egli, T., and Eggen, R. (1998) Molecular-based methods can contribute to assessments

of toxicological risks and bioremediation strategies. *J Microbiol Meth* 32, 107–119.

5. Mc Arthur, J.V. (2006) Microbial ecology. An evolutionary approach. Elsevier, Academic Press Publ., Amsterdam, New York.
6. Maes, A., Van Raemdonck, H., Smith, K., Ossieur, W., Lebbe, L., and Verstraete, W. (2006) Transport and activity of *Desulfitobacterium dichloroeliminans* strain DCA1 during bioaugmentation of 1,2-DCA-contaminated groundwater. *Environ Sci Technol* 40, 5544–5552.
7. De Wildeman, S., Neumann, A., Diekert, G., and Verstraete, W. (2003) Growth-substrate dependent dechlorination of 1,2-dichloroethane by a homoacetogenic bacterium. *Biodegradation* 14, 241–247.
8. Marzorati, M., de Ferra, F., Van Raemdonck, H., Borin, S., Allifranchini, E., Carpani, G., et al. (2007) A novel reductive dehalogenase, identified in a contaminated groundwater enrichment culture and in *Desulfitobacterium dichloroeliminans* strain DCA1, is linked to dehalogenation of 1,2-dichloroethane. *Appl Environ Microbiol* 73, 2990–2999.
9. Regeard, C., Maillard, J., Dufraigne, C., Deschavanne, P., and Holliger, C. (2005) Indications for acquisition of reductive dehalogenase genes through horizontal gene transfer by *Dehalococcoides ethenogenes* strain 195. *Appl Environ Microbiol* 71, 2955–2961.
10. Murray, A.E., Preston, C.M., Massana, R., Taylor, L.T., Blakis, A., Wu, K., et al. (1998) Seasonal and spatial variability of bacterial and archaeal assemblages in the coastal waters near Anvers Island, Antarctica. *Appl Environ Microbiol* 64, 2585–2595.
11. Hammer, Ø., Harper, D.A.T., and Ryan, P.D. (2001) PAST: paleontological statistics software package for education and data analysis. *Palaeontol. Electronica 4. [Online.]* http://palaeo-electronica.org/2001_1/past/issue1_01.htm.
12. Edwards, E., McEwan, R., Travis, A., and Wallance, R. (2004) 16s DNA library-based analysis of ruminal bacterial diversity. *Antonie Leeuwenhoek* 86, 263–281.
13. Maillard, J., Regeard, C., and Holliger, C. (2005) Isolation and characterization of Tn-Dha1, a transposon containing the tetrachloroethene reductive dehalogenase of *Desulfitobacterium hafniense* strain TCE1. *Environ Microbiol* 7, 107–117.
14. Marzorati, M., Borin, S., Brusetti, L., Daffonchio, D., Marsilli, C., Carpani, G., et al. (2006) Response of 1,2-dichloroethane-adapted microbial communities to ex-situ biostimulation of polluted groundwater. *Biodegradation* 17, 143–158.

Chapter 16

Methods for the Isolation of Genes Encoding Novel PHB Cycle Enzymes from Complex Microbial Communities

Ricardo F. Nordeste, Maria A. Trainer, and Trevor C. Charles

Abstract

Development of different PHAs as alternatives to petrochemically derived plastics can be facilitated by mining metagenomic libraries for diverse PHA cycle genes that might be useful for synthesis of bioplastics. The specific phenotypes associated with mutations of the PHA synthesis pathway genes in *Sinorhizobium meliloti* allows for the use of powerful selection and screening tools to identify complementing novel PHA synthesis genes. Identification of novel genes through their function rather than sequence facilitates finding functional proteins that may otherwise have been excluded through sequence-only screening methodology. We present here methods that we have developed for the isolation of clones expressing novel PHA metabolism genes from metagenomic libraries.

Key words: PHA/PHB cycle, *Sinorhizobium meliloti*, Microbial community gene libraries, Phenotypic complementation

1. Introduction

It is now well recognized that the majority of microbial community members are not represented in the culturable fraction. Metagenomic analyses of complex microbial communities necessitate a multifaceted approach that involves both sequence-based analyses and phenotypic selection. The use of phenotypic selection techniques represents a powerful tool for the isolation of truly novel genes that would not otherwise be identified on the basis of sequence alone (1).

Polyhydroxyalkanoates (PHA) represent a class of microbial polyesters composed of hydroxyacyl monomers, of which polyhydroxybutyrate (PHB) is the best-studied member (2). In the bacterial cell, PHAs are synthesized as cytoplasmically localized,

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668,
DOI 10.1007/978-1-60761-823-2_16, © Springer Science+Business Media, LLC 2010

electron-transparent granules, under conditions of abundant carbon when growth is limited by the availability of another key nutrient (3). The elastomeric and biodegradative properties of PHAs have generated considerable interest as potentially economically competitive, environmentally benign replacements to petrochemically derived plastics (4). Additionally, the potential use of PHAs in the medical field as materials for biocompatible surgical implants is promising (5, 6). Indeed, the potential commercial value of PHAs has generated the interest that has driven much of the research in this field.

The mechanical and physical properties of PHAs vary depending on the nature of the hydroxyacyl monomers (7). A range of properties, including melting point, elasticity, and tensile strength, may be altered by changing the composition of the monomer subunits (reviewed in ref. 2). The type of PHA synthesized by a given bacterial species depends on a multitude of factors, including the precursors provided to the polymerase enzyme responsible for construction of the polymer as well as the nature of the polymerase itself. The regulation of PHA accumulation and degradation has also been shown to involve a class of proteins called phasins (8). Phasins are involved in PHA granule formation, specifically in determining the size and number of PHA granules (8).

The cellular role of PHB, although not fully understood, is known to extend further than simply acting as an intracellular carbon store that can be mobilized to provide a bacterium with a competitive advantage over other soil microbes. PHAs have been shown to protect the cell from a wide range of stresses including heat shock, UV irradiation, exposure to oxidizing agents, and osmotic shock (9). PHB metabolism is also tightly linked to the redox state of the cell; previous studies have shown that in some bacteria, large quantities of PHB are accumulated under conditions of oxygen limitation (10–12). Furthermore, it has been suggested that PHB synthesis may act as an alternative electron acceptor under conditions of oxygen limitation; NAD(P)H is channeled into PHB formation to relieve inhibition of isocitrate dehydrogenase and citrate synthase, to allow continued operation of the TCA cycle (10, 13, 14).

Figure 1 shows the PHB cycle elucidated in *Sinorhizobium meliloti*, the nitrogen-fixing symbiont of alfalfa. Mutants of several of these PHB cycle enzymes have demonstrated interesting and informative phenotypes (15–18) that may be exploited as easy selection methods for the recovery of complementing clones from metagenomic libraries. Such clones could contain genes with interesting or valuable properties. Phenotypic screens that have been previously described include exopolysaccharide synthesis (19), staining with lipophilic dyes (20), fatty-acid detoxification (21), and nutritional auxotrophy (18, 22).

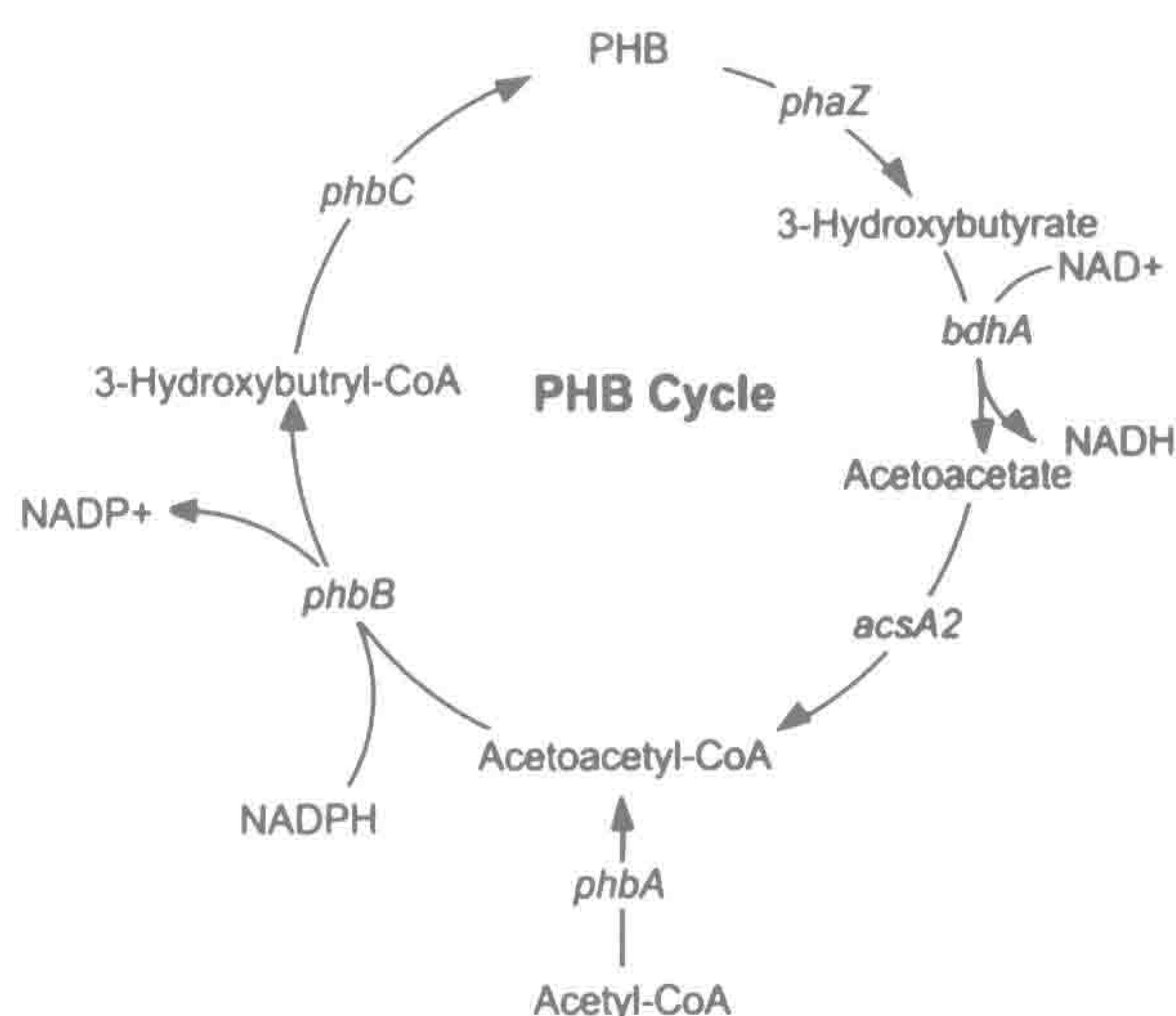


Fig. 1. PHB cycle of *S. meliloti*.

2. Materials

2.1. Bacterial Growth Media

1. Luria–Bertani Medium (LB) (23): 5 g yeast extract, 10 g Tryptone, 5 g NaCl, 1 L dH_2O, (15 g agar).
2. Tryptone yeast extract medium (TY) (24): 5 g Tryptone, 3 g yeast extract, 0.5 g $CaCl_2$, 1 L dH_2O, (15 g agar).
3. Modified M9 medium for *Rhizobium* (see Note 1) (25): 7 g Na_2HPO_4, 3 g KH_2PO_4, 1 g NH_4Cl, 1 g NaCl, (15 g agar). This is autoclaved, cooled to 55°C, and the following are added sterilely: 1 mL 0.5 M $MgSO_4$ and 0.1 mL 1 M $CaCl_2$.
4. *Rhizobium* minimal medium (RMM) (26): Solutions A, B, C, and D are prepared and sterilized separately. RMM is made by adding 1% (v/v) each of RMM A and RMM B and 0.1% (v/v) each of RMM C and RMM D.
   (a) RMM A: 145 g KH_2PO_4, 205 g K_2HPO_4, 15 g NaCl, 50 g NH_4NO_3, 1 L dH_2O.
   (b) RMM B: 50 g $MgSO_4 \cdot 7H_2O$, 1 L dH_2O.
   (c) RMM C: 10 g $CaCl_2 \cdot 2H_2O$, 1 L dH_2O.
   (d) RMM D: 123.3 g $MgSO_4 \cdot 7H_2O$, 87 g K_2SO_4, 0.247 g H_3BO_3, 0.1 g $CuSO_4 \cdot 5H_2O$, 0.338 g $MnSO_4 \cdot H_2O$, 0.288 g $ZnSO_4 \cdot 7H_2O$, 0.056 g $CoSO_4 \cdot 7H_2O$, 0.048 g $Na_2MoO_4 \cdot 2H_2O$, 1 L dH_2O.

5. Yeast mannitol medium (YM): 0.4 g yeast extract, 10 g mannitol, 0.5 g K_2HPO_4, 0.2 g $MgSO_4 \cdot 7H_2O$, 0.1 g NaCl, 1 L dH_2O, pH to 7.0, (18 g agar).
6. Yeast mannitol medium with Nile Red (YM-NR): 0.4 g yeast extract, 10 g mannitol, 0.5 g K_2HPO_4, 0.2 g $MgSO_4 \cdot 7H_2O$, 0.1 g NaCl, 1 L dH_2O, pH to 7.0, (18 g agar), 0.5 μg/mL Nile Red.

2.2. Molecular Biology Reagents

1. Ultra Clean Soil DNA Kit (Mo Bio Laboratories, Inc., Carlsbad, CA).
2. Gigapack III XL Lambda packaging extract (Stratagene).
3. EpiCentre™ EZ-Tn5 Insertion Kit (EpiCentre).
4. Small-scale plasmid preparation solution I: 50 mM glucose, 25 mM Tris–HCl pH 8.0, 10 mM EDTA pH 8.0.
5. Small-scale plasmid preparation solution II: 0.2 N NaOH, 1% SDS.
6. Small-scale plasmid preparation solution III: 60 mL 5 M potassium acetate, 11.5 mL glacial acetic acid, 28.5 mL dH_2O, store at 4°C.
7. $T_{10}E_{25}$: 10 mM Tris–HCl pH 8.0, 25 mM EDTA pH 8.0.
8. $T_{10}E_1$: 10 mM Tris–HCl pH 8.0, 1 mM EDTA pH 8.0.
9. TAE buffer: 40 mM Tris-acetate, 1 mM EDTA.
10. 40× TAE buffer: 242 g Tris base, 57.1 mL glacial acetic acid, 100 mL 0.5 M EDTA, pH 8.0.
11. Phage dilution buffer: 10 mM Tris–HCl pH 8.3, 100 mM NaCl, 10 mM $MgCl_2$.

3. Methods

3.1. Bacterial Growth and Storage Conditions

1. *Escherichia coli* strains are routinely grown at 37°C using Luria–Bertani (LB) medium (23). *S. meliloti* strains are routinely cultured at 30°C in either LB (23) or TY (24) medium. When *S. meliloti* is grown in modified M9 (25) or RMM (26), the medium is supplemented with 15 mM glucose, D-3-hydroxybutyrate (D3HB), L-3-hydroxybutyrate (L3HB), DL-3-hydroxybutyrate (DLHB), acetoacetate (AA), or acetate as the carbon source. For growth under high-carbon conditions, *S. meliloti* is cultured in yeast mannitol (YM) medium.
2. Antibiotics are used in the growth medium where appropriate. Concentrations for *E. coli* are as follows: ampicillin 100 μg/mL, chloramphenicol 25 μg/mL, gentamycin 10 μg/mL, kanamycin 25 μg/mL, nalidixic acid 5 μg/mL,

tetracycline 10 μg/mL. Concentrations for *S. meliloti* are as follows: gentamycin 75 μg/mL, neomycin 200 μg/mL, spectinomycin 100 μg/mL, streptomycin 200 μg/mL, tetracycline 10 μg/mL, trimethoprim 400 μg/mL.

3. All bacterial cultures are stored at –70°C in glass cryovials containing 7% dimethyl sulfoxide (DMSO).

3.2. Construction of Metagenomic Libraries from Soil

1. Total DNA from soil samples is isolated using Ultra Clean Soil DNA Kit, followed by a phenol–chloroform extraction with two additional chloroform extractions.
2. The DNA is incubated with RNase for 20 min at 37°C to remove all traces of RNA.
3. The DNA is cleaned by isopropanol precipitation, washed twice with 75% EtOH, and resuspended in $T_{10}E_1$ buffer.
4. Cosmid libraries are constructed by cloning fragments from BamHI partial digests (see Note 3) into the BamHI site of the IncP Tc^R plasmid pRK7813 (27), followed by packaging with Gigapack III XL Lambda packaging extract and transduction of *E. coli* HB101 (28).
5. Tc^R colonies are selected, and representative library clones are analyzed by restriction digest.
6. Colonies are pooled and subcultured. The resultant libraries are maintained at –70°C as aliquots in LB containing 7% DMSO.

3.3. Transfer of a Metagenomic Library into S. meliloti by Triparental Conjugation

Conjugation is typically performed by triparental mating between *E. coli* donors carrying the metagenomic library, an *E. coli* strain carrying a helper plasmid, and an *S. meliloti* recipient.

1. Wash all strains in 0.85% NaCl to remove antibiotics.
2. Combine 1 mL saturated broth culture of the *S. meliloti* recipient with 500 μL each of the *E. coli* donor and helper strains.
3. Recover cells by centrifugation, resuspend in 20 μL 0.85% NaCl, and spot onto a nonselective LB or TY plate.
4. Incubate overnight at 30°C.
5. Resuspend mating spot in 1 mL 0.85% NaCl.
6. Prepare serial dilutions of the resuspended mating spot in 0.85% NaCl; plate 100 μL of the appropriate dilutions onto selective medium and incubate overnight at 30°C.

3.4. Transfer of Putative Complementing Clones from S. meliloti into E. coli

1. Wash all strains in 0.85% NaCl to remove antibiotics.
2. Combine 1 mL saturated broth culture of the *S. meliloti* recipient with 500 μL each of the *E. coli* donor and helper strains.

3. Recover cells by centrifugation, resuspend in 20 µL 0.85% NaCl, and spot onto a nonselective LB or TY plate.
4. Incubate overnight at 30°C.
5. Resuspend mating spot in 1 mL 0.85% NaCl.
6. Prepare serial dilutions of the resuspended mating spot in 0.85% NaCl; plate 100 µL of the appropriate dilutions onto selective medium and incubate overnight at 37°C.
7. Restreak transconjugants onto selective medium and isolate the cosmid DNA by standard plasmid isolation techniques (23).

3.5. Genetics and Molecular Biology

In vitro Tn5 mutagenesis of plasmid DNA is used to generate transposon mutations that facilitate the subsequent determination of plasmid DNA sequence. This mutagenesis is performed using the EpiCentre™ EZ-Tn5 Insertion Kit as per the manufacturer's instructions.

3.6. PHA Accumulation

PHB content is determined using a modified version of the colorimetric assay developed by Law and Slepecky (29). This assay is based on the hydrolysis of PHB and subsequent conversion of the monomer to crotonic acid by concentrated H_2SO_4. Crotonic acid has an absorption maximum at 235 nm. The amount of crotonic acid can be used to determine PHB content of the initial sample. PHB content is expressed as a percentage of total cellular dry mass. Do not use any plasticware in this protocol (see Note 2).

1. Pellet cells in screw-capped Pyrex centrifuge tubes at 7,000 rcf in an IEC 21000R centrifuge with a 7685c rotor (or equivalent) for 10 min.
2. Wash cell pellet in dH_2O and pellet again.
3. Resuspend pellet in 2.0 mL of 5.25% NaOCl and incubate at 37°C for 1 h to allow complete cell lysis to occur.
4. Pellet samples at 7,000 rcf for 15 min and wash in 5 mL dH_2O followed by 5 mL EtOH and finally 5 mL acetone.
5. The pellet, which should be white in color, should be allowed to dry before the PHB is extracted.
6. PHB is extracted by the addition of 10 mL of cold chloroform. The tubes should capped, vortexed, and transferred to a boiling water bath. The tubes are removed from the water bath and vortexed every 1–2 min for 10 min before cooling to room temperature. The PHB should now be dissolved in the chloroform.
7. Once cool, the tubes are vortexed again, and 1 mL is removed and transferred to a glass test tube.
8. The chloroform should be allowed to evaporate completely at room temperature (see Note 4) (should take 24–48 h) before addition of 10 mL concentrated H_2SO_4.

9. The tubes are then capped with marbles (to prevent entry of water and pressure build up) and transferred to a boiling water bath for 10 min, after which time they are removed and allowed to cool to room temperature.
10. After mixing well by vortex, OD from 220 to 280 nm is measured, and PHB is quantified by comparison to data generated by a standard curve (see Note 5).

PHA deposits may also be visualized by transmission electron microscopy:

1. Samples are prepared from 100 mL stationary phase YM cultures.
2. Cells are harvested by centrifugation, suspended in phosphate buffer (pH 6), and collected by centrifugation.
3. The cells are then suspended in 1 mL of 2.5% glutaraldehyde in phosphate buffer, and kept at 4°C for 1 h, followed by three series of centrifugation and resuspension in 1 mL of phosphate buffer.
4. The washed cells are suspended in 1 mL of 0.5% OsO_4 in phosphate buffer and kept at room temperature for 16 h, and then diluted to 8 mL in phosphate buffer.
5. The cells are collected by centrifugation and resuspended in 2% agar, a drop of which is then allowed to harden on a microscope slide.
6. The agar-suspended cells are then dehydrated in a series from 50 to 100% acetone washes, embedded in eponaraldite, sectioned at a thickness of 60–90 nm on a Reichert Ultracut E ultramicrotome (or equivalent), stained with uranyl acetate and lead citrate, and examined on a Philips CM10 transmission electron microscope (or equivalent) using an accelerating voltage of 60 kV.

3.7. Screening of Metagenomic Libraries for PHB Synthesis Clones

The distinct, nonmucoid colony morphology exhibited by PHB synthesis mutants of *S. meliloti* provides a novel and powerful screen for complementation by PHB synthesis genes (19, 28). This screen may be further enhanced by the inclusion of 0.5 μg/mL Nile Red into the YM agar (YM-NR); PHB-synthesizing colonies will stain pink, while nonsynthesizing colonies remain unpigmented (see Note 6).

3.7.1. Complementation

1. The metagenomic libraries are introduced en masse into S. meliloti Rm11476 by triparental conjugation.
2. Transconjugants are selected on YM-NR agar supplemented with an appropriate antibiotic – Tc for libraries constructed in pRK7813; Nm for selection in Rm11476 – and the resultant colonies are screened for pink coloration and mucoidy.

3. Clones from pink, mucoid colonies are transferred to *E. coli* DH5α by triparental conjugation.
4. The complementing cosmids are then reintroduced into *S. meliloti phbC* mutants to confirm the associated colony and growth.
5. Cosmid DNA from complementing clones is analyzed to identify clones exhibiting unique restriction patterns.
6. PHB accumulation is confirmed in the transconjugants by PHB assay and transmission electron microscopy.

3.7.2. Sequence Analysis of Complementing Clones

1. Complementing clones exhibiting unique restriction patterns are subcloned into pBBR1MCS-5 (30); complementing clones are identified by complementation analysis on YM-NR.
2. The complementing region is localized by EZ-Tn-Kan-2 in vitro mutagenesis and subsequent subcloning steps; the DNA sequence of the complementing clones is facilitated by the transposon insertions.
3. The resultant DNA sequence is compared with other sequences by BLASTX analysis (31).

3.8. Utilization of Nutritional Auxotrophy to Facilitate the Isolation of PHB Cycle Genes from Metagenomic Libraries

Mutants of different PHB cycle genes exhibit nutritional auxotrophies that represent powerful selection tools for the isolation of complementing clones from metagenomic libraries. These are summarized in Table 1.

3.8.1. Complementation

1. The metagenomic libraries are introduced en masse into the appropriate *S. meliloti* mutant.
2. Transconjugants are selected on RMM or M9 agar supplemented with appropriate antibiotic to counterselect the *E. coli* donor and an appropriate carbon source.
3. Isolated clones are screened for the presence of a cosmid by patching onto LB or TY medium containing the appropriate antibiotic (Tc for libraries constructed in pRK7813).
4. Cosmids from the resulting colonies are transferred to *E. coli* DH5α by triparental conjugation.
5. Complementing cosmids are then reintroduced into the appropriate *S. meliloti* PHB cycle mutant to confirm complementation on the appropriate carbon source.
6. Cosmid DNA from complementing clones is analyzed to identify clones exhibiting unique restriction patterns.

Table 1
Nutritional auxotrophies and colony phenotypes of *S. meliloti* PHB cycle mutants

ORF	Auxotrophy	Nile Red	Mucoidy	Reference
WT	None	+	+	(32)
phbA	No growth on acetoacetate	–	–	(unpublished)
phbB	Poor growth on D-3-hydroxybutyrate and acetoacetate	–	–	(19)
phbC	Poor growth on D-3-hydroxybutyrate and acetoacetate	–	–	(18, 33)
phaZ	None	+	++	(unpublished)
bdhA	No growth on D-3-hydroxybutyrate	+	+	(34)
acsA2	Poor growth on acetoacetate	+	+	(33)
phaP1	Slow growth on succinate	+	+	(35)
phaP2	Slow growth on succinate	+	+	(35)
phaP1/P2	Slow growth on succinate	–	++	(35)

3.8.2. Sequence Analysis of Complementing Clones

1. Complementing clones exhibiting unique restriction patterns are subcloned into pBBR1MCS-5 (30); complementing clones are identified by selection for growth on the appropriate carbon source.
2. The complementing region is localized by EZ-Tn-Kan-2 in vitro mutagenesis and subsequent subcloning steps; the DNA sequence of the complementing clones is facilitated by the transposon insertions.
3. The resultant DNA sequence is compared with other sequences by BLASTX analysis (31).

4. Notes

1. To facilitate the growth of rhizobial species, M9 medium is modified to include 0.25 mM $CaCl_2$, 1 mM $MgSO_4$, and 0.3 mg/L biotin.
2. Following the initial cell harvest, no plasticware should be used in the PHB extraction protocol; all glasswares used must be washed thoroughly in boiling chloroform and rinsed in EtOH prior to use, to remove any traces of plasticizers.
3. Partial digestion of genomic DNA is optimized by gradient digest. In a tube, 15 μL genomic DNA (15 ng/μL) is mixed

with 100 μL 10× digest buffer in a final volume of 500 μL. This mixture is incubated on ice for 30 min. The reaction mix is then aliquoted into 15 tubes (60 μL is added to the first tube; 30 μL is added to the remaining 14 tubes), and 5 units of the Sau3AI is added to the first tube. A concentration gradient is established by transferring 30 μL from the first tube into the second, mixing, and then transferring 30 μL from tube 2 into tube 3 and so on. 30 μL is removed from the final tube and discarded. The reactions are incubated at 37°C for 30 min, and the reactions are stopped by the addition of 1 μL 0.5 M EDTA mixed with 6× loading dye. The digests are run on an agarose gel, and the enzyme concentration that gives fragments of approximately 25–50 kb is selected for subsequent use.

4. Evaporation of the chloroform can be expedited by gentle heating at 40°C.
5. A standard curve is obtained by assaying known quantities of PHB. Standard solutions are prepared from a 1 mg/mL PHB stock, made by adding 10 mg PHB to 10 mL cold chloroform and heating in a boiling water bath to dissolve, as above. From this, a 100 μg/mL stock is prepared. Aliquots of 0–100 μg PHB are transferred to test tubes, and the chloroform is allowed to evaporate before addition of 10 mL H_2SO_4 and processing, as described above.
6. The *exoY*::Tn5 mutant Rm7055 (36, 37), in which the extracellular polysaccharide succinoglycan is not produced, forms nonmucoid colonies that fluoresce brightly under UV illumination. Strain Rm11476, containing both *exoY*::Tn5 and *phbC*::Tn5-233 mutations, forms nonmucoid colonies that do not stain or fluoresce. This is the best genetic background for the detection of clones that complement for PHB accumulation, especially on densely populated plates.

References

1. Henne, A., Daniel, R., Schmitz, R.A., and Gottschalk, G. (1999) Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl Environ Microbiol* 65, 3901–3907.
2. Anderson, A.J. and Dawes, E.A. (1990) Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. *Microbiol Rev* 54, 450–472.
3. Zevenhuizen, L.P.T.M. (1981) Cellular glycogen, β-1,2-glucan, poly-3-hydroxybutyic acid and extracellular polysaccharides in fast-growing species of *Rhizobium*. *Antonie Van Leeuwenhoek* 47, 481–497.
4. Madison, L.L. and Huisman, G.W. (1999) Metabolic engineering of poly(3-hydroxyalkanoates): from DNA to plastic. *Microbiol Mol Biol Rev* 63, 21–53.
5. Shishatskaya, E.I., Voinova, O.N., Goreva, A.V., Mogilnaya, O.A., and Volova, T.G. (2008) Biocompatibility of polyhydroxybutyrate microspheres: *in vitro* and *in vivo* evaluation. *J Mater Sci Mater Med* 19, 2493–2502.
6. Shishatskaya, E.I., Volova, T.G., Puzyr, A.P., Mogilnaya, O.A., and Efremov, S.N. (2004) Tissue response to the implantation of biodegradable polyhydroxyalkanoate sutures. *J Mater Sci Mater Med* 15, 719–728.

7. Holmes, P.A. (1985) Applications of PHB – a microbially produced biodegradable thermosplastic. *Phys Technol* 16, 32–36.
8. Pötter, M. and Steinbüchel, A. (2005) Poly(3-hydroxybutyrate) granule-associated proteins: impacts on poly(3-hydroxybutyrate) synthesis and degradation. *Biomacromolecules* 6, 552–560.
9. Kadouri, D., Jurkevitch, E., and Okon, Y. (2003) Involvement of the reserve material poly-β-hydroxybutyrate in *Azospirillum brasilense* stress endurance and root colonization. *Appl Environ Microbiol* 69, 3244–3250.
10. Senior, P.J., Beech, G.A., Ritchie, G.A.F., and Dawes, E.A. (1972) The role of oxygen limitation in the formation of poly-3-hydroxybutyrate during batch and continuous culture of *Azotobacter beijerinckii*. *Biochem J* 128, 1193–1201.
11. Stam, H., van Verseveld, H.W., de Vries, W., and Stouhamer, A.H. (1986) Utilization of poly-β-hydroxybutyrate in free-living cultures of *Rhizobium* ORS571. *FEMS Microbiol Lett* 35, 215–220.
12. Stockdale, H., Ribbons, D.W., and Dawes, E.A. (1968) Occurrence of poly-3-hydroxybutyrate in the Azotobacteriaceae. *J Bacteriol* 95, 1798–1803.
13. Senior, P.J. and Dawes, E.A. (1971) Poly-3-hydroxybutyrate biosynthesis and the regulation of glucose metabolism in *Azotobacter beijinkereii*. *Biochem J* 125, 55–66.
14. Page, W.J. and Knosp, O. (1989) Hyperproduction of poly-3-hydroxybutyrate during exponential growth of *Azotobacter vinelandii* UWD. *Appl Environ Microbiol* 55, 1334–1339.
15. Aneja, P. and Charles, T.C. (1999) Poly-3-hydroxybutyrate degradation in *Rhizobium* (*Sinorhizobium*) *meliloti*: isolation and characterization of a gene encoding 3-hydroxybutyrate dehydrogenase. *J Bacteriol* 181, 849–857.
16. Aneja, P., Dziak, R., Cai, G.Q., and Charles, T.C. (2002) Identification of an acetoacetyl coenzyme-A synthetase-dependent pathway for utilization of L-(+)-3-hydroxybutyrate in *Sinorhizobium meliloti*. *J Bacteriol* 184, 1571–1577.
17. Charles, T.C., Cai, G.Q., and Aneja, P. (1997) Megaplasmid and chromosomal loci for the PHB degradation pathway in *Rhizobium* (*Sinorhizobium*) *meliloti*. *Genetics* 146, 1211–1220.
18. Willis, L.B. and Walker, G.C. (1998) The phbC (poly-β-hydroxybutyrate synthase) gene of *Rhizobium* (*Sinorhizobium*) meliloti and characterization of phbC mutants. *Can J Microbiol* 44, 554–564.
19. Aneja, P., Dai, M., Lacorre, D.A., Pillon, B., and Charles, T.C. (2004) Heterologous complementation of the exopolysaccharide synthesis and carbon utilization phenotypes of *Sinorhizobium meliloti* Rm1021 polyhydroxyalkanoate synthesis mutants. *FEMS Microbiol Lett* 239, 277–283.
20. Ostle, A.G. and Holt, J.G. (1982) Nile blue as a fluorescent stain for poly-β-hydroxybutyrate. *Appl Environ Microbiol* 44, 238–241.
21. Kranz, R.G., Gabbert, K.K., and Madigan, M.T. (1997) Positive selection systems for discovery of novel polyester biosynthesis genes based on fatty acid detoxification. *Appl Environ Microbiol* 63, 3010–3013.
22. Povolo, S., Tombolini, R., Morea, A., Anderson, A.J., Casella, S., and Nuti, M.P. (1994) Isolation and characterization of mutants of *Rhizobium meliloti* unable to synthesize poly-3-hydroxybutyrate (PHB). *Can J Microbiol* 40, 823–829.
23. Sambrook, J. and Russell, D.W. (2001) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Press, New York.
24. Beringer, J.E. (1974) R factor transfer in *Rhizobium leguminosarum*. *J Gen Microbiol* 84, 188–198.
25. Miller, J.H. (1972) Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
26. Dowling, D.N., Samrey, U., Stanley, J., and Broughton, W.J. (1987) Cloning of *Rhizobium leguminosarum* genes for competitive nodulation blocking on peas. *J Bacteriol* 169, 1345–1348.
27. Jones, J.D. and Gutterson, N. (1987) An efficient mobilizable cosmid vector, pRK7813, and its use in a rapid method for marker exchange in *Pseudomonas fluorescens* strain HV37a. *Gene* 61, 299–306.
28. Wang, C., Meek, D.J., Panchal, P., Boruvka, N., Archibald, F.S., Driscoll, B.T., et al. (2006) Isolation of poly-3-hydroxbutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants. *Appl Environ Microbiol* 72, 384–391.
29. Law, J. and Slepecky, R. (1961) Assay of poly-3-hydroxybutyric acid. *J Bacteriol* 82, 33–36.
30. Kovach, M.E., Elzer, P.H., Hill, D.S., Robertson, G.T., Farris, M.A., Roop, R.M., et al. (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* 166, 175–176.

31. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389–3402.

32. Meade, H.M., Long, S.R., Ruvkun, G.B., Brown, S.E., and Ausubel, F.M. (1982) Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn5 mutagenesis. *J Bacteriol* 149, 114–122.

33. Cai, G., Driscoll, B.T., and Charles, T.C. (2000) Requirement for the enzymes acetoacetyl coenzyme-A synthetase and poly-3-hydroxybutyrate (PHB) synthase for growth of *Sinorhizobium meliloti* on PHB cycle intermediates. *J Bacteriol* 182, 2113–2118.

34. Aneja, P., Zachertowska, A., and Charles, T.C. (2005) Comparison of the symbiotic and competition phenotypes of *Sinorhizobium meliloti* PHB synthesis and degradation pathway mutants. *Can J Microbiol* 51, 599–604.

35. Wang, C.X., Sheng, X.Y., Equi, R.C., Trainer, M.A., Charles, T.C., and Sobral, B.W.S. (2007) Influence of the poly-3-hydroxybutyrate (PHB) granule-associated proteins (PhaP1 and PhaP2) on PHB accumulation and symbiotic nitrogen fixation in *Sinorhizobium meliloti* Rm1021. *J Bacteriol* 189, 9050–9056.

36. Leigh, J.A., Signer, E.R., and Walker, G.C. (1985) Exopolysaccharide-deficient mutants of *Rhizobium meliloti* that form ineffective nodules. *Proc Natl Acad Sci U S A* 82, 6231–6235.

37. Miller-Williams, M., Loewen, P.C., and Oresnik, I.J. (2006) Isolation of salt-sensitive mutants of *Sinorhizobium meliloti* strain Rm1021. *Microbiology* 152, 2049–2059.

Chapter 17

Metagenomic Approaches to Identify and Isolate Bioactive Natural Products from Microbiota of Marine Sponges

Cristian Gurgui and Jörn Piel

Abstract

Many marine sponges harbor massive consortia of symbiotic bacteria belonging to diverse phyla. Sponges are also an unusually rich source of biologically active natural products, and evidence is accumulating that these compounds might often be synthesized by the symbionts. Since the study of sponge-associated bacteria is generally hampered by very low cultivation rates, cultivation-independent, metagenomic methods have recently been applied to sponges. These methods allow for the isolation of biosynthetic gene clusters that can ultimately be exploited to develop sustainable natural product sources by heterologous expression. However, general challenges encountered in sponge metagenomic research are the poor quality of the isolated DNA with respect to size and yield, the difficulty to identify genes of interest among numerous homologs, insufficient clone numbers in metagenomic libraries, and time-consuming screening procedures to identify and isolate rare positive clones. Here, we give an overview of methods that address these problems and can be used to streamline isolation of biosynthetic and other genes of interest.

Key words: Metagenomics, Marine sponges, Sponge symbionts, Natural products, Polyketide synthases

1. Introduction

Sponges are the simplest multicellular animals and have neither true tissues nor organs (1). Their morphology caused ancient scientists like Aristotle and Pliny to consider them as plants (2). It was John Ellis in 1765 who proved that they are actually animals (3). As filter feeders, sponges can process almost 20,000 times their volume in water per day (4), thus accumulating bacteria, viruses, fungi, and other particulate matter present in sea habitats. With an age of more than 635 million years, the phylum Porifera (sponges) is particularly successful (5), which might be, in part,

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668,
DOI 10.1007/978-1-60761-823-2_17,

due to an unusually rich defensive chemistry present in many species (6). With often excellent pharmacological properties, sponge-derived natural products are good candidates for the development of drugs with antibiotic, antiviral, antitumor, and anti-inflammatory activities. Since many of these compounds structurally resemble complex polyketides and nonribosomal peptides from bacteria (7, 8), it is likely that the true producers are symbiotic. This hypothesis is supported by cell separation studies that localized compounds to prokaryotes (9) and by the isolation of bacterial genes involved in production of "sponge" polyketides from the total animal DNA (10). The study of the biosynthetic capabilities of symbionts is gaining considerable attention, since it could provide a fundamentally new perspective to marine pharmacology. Currently, drug development is delayed or even inhibited since often ecologically inacceptable amounts of sponge material would be required for pharmacological evaluation. In addition, many compounds exhibit highly complex structures that prohibit total chemical synthesis on a large scale (11, 12). In case of a bacterial production, alternative supplies could be generated by cultivation or heterologous expression of biosynthetic genes. At present, the latter approach appears to be more generally useful, since to date only very few true sponge symbionts have been successfully cultivated (13). This strategy largely employs metagenomic techniques (14) and relies on the analysis of complex DNA mixtures isolated from the entire sample or from enriched preparations. The DNA is cloned in a large-insert vector and used to construct an environmental DNA library in a culturable bacterium, most commonly *Escherichia coli*. By using appropriate PCR primers or hybridization probes, the biosynthetic genes of a natural product can be identified and isolated for further study and heterologous expression (15).

2. Materials

2.1. Metagenomic DNA Isolation from Sponges

1. RNAlater (Qiagen) or 100% ethanol for storage of sponge samples (see Note 1).
2. Sponge lysis buffer: 8 M urea, 2% sarkosyl (Sigma–Aldrich Co., St. Louis, USA), 1 M NaCl, 50 mM EDTA, and 50 mM Tris–HCl, pH 7.5. Prepare fresh. The buffer is either autoclaved or sterile-filtered and stored at room temperature.
3. Phenol/$CHCl_3$/isoamyl alcohol (25:24:1 v:v:v): store at 4°C.
4. $CHCl_3$: store at room temperature.
5. Cetyl trimethylammonium bromide (CTAB; Sigma–Aldrich Co., St. Louis, USA): a stock solution of 10% (w/v) is prepared and stored at room temperature (see Note 2).

6. 3 M sodium acetate, pH 7: store at room temperature.
7. 70% Ethanol: store at –20°C.
8. 10 mM Tris–HCl buffer, pH 8.5: store at room temperature.
9. Agarose: store at room temperature. Working solution is 1% (w/v).
10. Ethidium bromide 1% stock solution: store at 4°C. Use 1 μL for each 100 mL of 1% agarose solution.

2.2. Metagenomic DNA Library Construction

1. CopyControl™ Fosmid Library Production Kit (EPICENTRE Biotechnologies, Madison, USA): Store according to the manufacturer's instructions.
2. Low-melting point (LMP) agarose: Store at room temperature. Working solution is usually 1% (w/v).
3. TAE buffer for electrophoresis: A 50× 1 L stock solution contains 242 g Tris base, 57.1 mL acetic acid, and 100 mL 0.5 M EDTA. Adjust pH to 8.5 with KOH and store at room temperature. Electrophoresis running buffer is 1× TAE.
4. Gel electrophoresis system (Bio-Rad Laboratories, Hercules, USA).
5. Gel documentation system (Syngene, Cambridge, UK).
6. 100% Ethanol: Store an aliquot at room temperature and one at –20°C.
7. 3 M sodium acetate, pH 7: Store at room temperature.
8. 1 M EDTA: Store at room temperature.
9. 70% Ethanol: Store at –20°C.
10. 10 mM Tris–HCl buffer pH 8.5: Store at room temperature.
11. Isopropanol: Store at room temperature.
12. 3 M sodium acetate, pH 5.2: Store at room temperature.
13. T4 DNA ligase (New England Biolabs, Ipswich, USA): Store at –20°C.
14. Thermomixer (Eppendorf, Hamburg, Germany) with a temperature control range 13°C below ambient to 99°C.
15. LB broth medium (recipe for 1 L): 10 g Bacto-tryptone, 5 g yeast extract, and 10 g NaCl (pH 7.5). The medium is supplemented with $MgSO_4$ to a final concentration of 10 mM. Store at room temperature.
16. Phage dilution buffer (PDB): 10 mM Tris–HCl, pH 8.3, 100 mM NaCl, and 10 mM $MgCl_2$. Store at room temperature.
17. LB broth agar (same as for LB broth medium, except that 15 g agar is also added to 1 L of medium).
18. Chloramphenicol: prepare a stock solution of 25 mg/mL in 100% ethanol. Store in aliquots at –20°C.

For the DNA electroelution protocol, the following additional materials are needed:

19. Dialysis tubing (ZelluTrans/V Serie; MWCO: 1,000; Width: 18 mm; Wall thickness: 23 µm; vol/cm: 1.04 mL; Carl Roth GmbH + Co. KG, Karlsruhe, Germany) and clips.

2.3. Metagenomic DNA Library Plating and Screening by PCR

1. SeaPrep agarose (Lonza Group Ltd, Basel, Switzerland). Semiliquid medium is prepared by mixing 5 g of SeaPrep agarose with 1 L of LB broth medium (see Note 3). Store at room temperature.
2. LB broth medium.
3. 2 mL screw-cap vials.
4. Chloramphenicol stock solution (25 mg/mL, prepared in 100% ethanol).
5. 50% Glycerol. Store at room temperature.
6. Taq DNA polymerase with ThermoPol buffer (New England Biolabs, Ipswich, USA), dNTPs (Invitrogen Corporation, Carlsbad, USA), BSA (New England Biolabs, Ipswich, USA). Store at –20°C.

3. Methods

It is critical that the sponge DNA isolation protocol yields high-quality metagenomic DNA: high molecular weight, high yield, and integrity (no shearing). First, a low molecular weight DNA (below 30 kbp) would not make a viable lambda phage able to infect bacteria. Second, because during the following steps after isolation (e.g., size selection and end repair) the DNA recovery is usually below 50%, it is important that the DNA isolation is highly efficient. Approximately 1 g of wet sponge (stored at 4°C or –20°C either in RNAlater or in 100% ethanol) should yield sufficient amounts of DNA for one library production trial. This can, of course, vary to a large extent with the sponge species, depending also on the amount of sponge-associated bacteria. Often the isolated DNA appears to be very good in size and concentration but still contains contaminations that inhibit PCR and library production steps. Further gel purification via LMP agarose produces high-quality DNA. Most sponges have a high polysaccharide content that might interfere with DNA handling and with subsequent enzymatic reactions (e.g., end repair, ligation). A cationic detergent, CTAB, is used at concentrations between 0.5 and 1% to remove polysaccharides. It is important to maintain a NaCl concentration in the lysis buffer above 0.5 M, or a CTAB–DNA precipitate will form (16). Heating at 65°C will be necessary to dissolve the 10% CTAB solution, and the stock should be reheated

each time before use to reduce viscosity. With some sponge species, further optimization is needed (see Note 4).

It is important to always elute DNA with Tris–HCl buffer and not with Tris–EDTA buffer, since the latter contains EDTA, which may inhibit or lower the efficiency of subsequent enzymatic reactions (e.g., end repair, ligation). In contrast to the protocol provided by the manufacturer of the fosmid library kit, the LMP size selection of the metagenomic DNA should be done before blunt ending (or end repair), since it avoids additional shearing of the blunt-ended DNA, thereby improving clone yields. In addition, the electrophoresis step removes contaminants that might interfere with enzymes used in the end-repair reaction. Blunt ending is a very critical step. The enzyme mixture used for performing this step contains T4 DNA polymerase, which exhibits both a template-directed 5′→3′ DNA polymerase activity and 3′→5′ exonuclease activity. Therefore, the incubation period should not be extended over the specified 45 min, and the reaction should be stopped accordingly (see protocol).

It was also found that the cloning efficiency is much higher when T4 DNA ligase (New England Biolabs) is used in an overnight reaction at 16°C. This enzyme worked better in our hands than the one provided in the Epicentre kit. During all steps when working with lambda phage packaging extracts, pipette tips with larger opening are used. This reduces the shearing forces exerted on them during all pipetting steps (also during titering of the library). Since the library is plated within 24 h, no chloroform is added to the packaging reaction at the end of the incubation time, but only PDB. For the preparation of large-sized libraries, it was found crucial that all steps from DNA isolation to library plating were done without freezing and thawing samples.

Timetable

Day 1: DNA isolation and DNA size selection on LMP agarose gel.
Day 2: DNA gel extraction blunt ending and ligation (overnight).
Day 3: Packaging reaction library titering.
Day 4: Plating of the library.

Starting at day 5: Library screening (takes about 4–5 days for a 400,000 member library).

An efficient library screening protocol was developed in our laboratory (17), which includes distributing the library into 2 mL screw-cap vials containing a semiliquid medium that allows the clones to grow spatially. There are four main advantages of distributing the library in such a semiliquid medium and not plating it on plates: space for cultivation and storage is reduced, the library can be stored immediately at −80°C after simple

addition of glycerol, the risk of cross-contamination is lower, and libraries can be screened very fast and in an economic way.

3.1. Metagenomic DNA Isolation from Sponges

1. Depending on the sponge species, this isolation procedure should usually yield between 5 and 10 µg total DNA, starting from 1 g of wet tissue.
2. Sponge samples should be stored at 4 to –80°C in RNAlater or 100% ethanol immediately after collection (see Note 1).
3. The sponge tissue (ca. 1 g wet weight) is cut off with a scalpel in a way to include both surface and inner parts.
4. The sponge material is ground to a fine powder under liquid nitrogen using a prechilled mortar and pestle. The resulting powder is quickly transferred to a 50 mL falcon tube containing 10 mL sponge lysis buffer. Use 1 mL of lysis buffer for each 100 mg of sponge tissue.
5. Mix gently so that the sponge particles are resuspended well and evenly distributed. Incubate the mixture for 10–20 min at 60°C with gentle mixing every 5 min to avoid formation of large sponge clumps (see Note 5).
6. Extract several times until bottom phase is clear of any debris, with 10 mL phenol/$CHCl_3$/isoamyl alcohol (25:24:1, v:v:v). Handle the mixture very gently without vigorous shaking or vortexing.
7. After each extraction, separate phases by centrifugation (see Note 6) and take the upper aqueous phase.
8. Take the aqueous phase and extract with the same volume of chloroform (this step also removes the phenol traces). Centrifuge again.
9. Save the upper phase and precipitate polysaccharides by adding CTAB to 0.5–1% final concentration (see Notes 2 and 4). Mix gently and incubate at 60°C for 10–15 min. A whitish precipitate forms upon CTAB addition.
10. Centrifuge for 15 min to remove the precipitated polysaccharides and save supernatant.
11. Precipitate DNA with two volumes of ice-cold 100% ethanol (stored at –20°C). Also add one-tenth volumes of 3 M sodium acetate (pH 7) to improve precipitation. Mix gently by slowly inverting the tube several times to mix the two layers. Prior to centrifugation, store at –20°C for 15–30 min to increase yield.
12. Centrifuge for 30 min and then carefully remove the supernatant without disturbing the DNA pellet.
13. Wash the pellet twice with 70% ethanol (this step also removes the CTAB traces, as CTAB is soluble in ethanol). Centrifuge for 10–15 min and remove the ethanol.

14. Air-dry DNA pellet for a few minutes (see Note 7) and then resuspend in 500 μL Tris–HCl buffer (10 mM, pH 8.5). If the pellet is very large and white in appearance, it is probably due to the large amounts of polysaccharides in the sponge sample, as well as the high urea concentration in the lysis buffer. In order to partially remove them, one can resuspend the entire pellet in about 1 mL of Tris–HCl buffer and then place the tube on ice for 10–15 min until crystals form. The clear liquid phase contains the highest fraction of DNA and is carefully removed and used during the following steps (see Note 8). However, we observed that DNA precipitation with isopropanol at room temperature reduces formation of such whitish big pellets. Alternate protocol for DNA resuspension (this ensures a better solubilization): remove most of the ethanol but leave a last drop on the pellet and add the Tris–HCl buffer. Mark the buffer level on the tube and add a few more drops of water. Speed-vac the suspension to remove any ethanol traces up to the marked level (see Note 9).
15. Run a 1% agarose gel to check for DNA quality (see Note 10). Estimate DNA size and concentration by loading different amounts of DNA sample and comparing them to a marker with known size and concentration (usually the 36 kb Fosmid Control Insert DNA included in the Epicentre Library Construction kit).
16. The DNA preparation should be very good in size and concentration; however, further gel purification via LMP is necessary to produce high-quality DNA.
17. Proceed with the LMP agarose gel (see below) by running half of the DNA amount (usually 2–5 μg) obtained after the isolation procedure. Store the other half at −20°C as a backup. An example of total DNA isolated from the sponge *Psammocinia* aff. *bulbosa* is shown in Fig. 1, lane 2.

3.2. Metagenomic DNA Library Construction

3.2.1. Size Selection of Metagenomic DNA on LMP Agarose

1. Prepare a 0.8–1% LMP agarose gel in 1× TAE buffer (see Note 11). Do not include ethidium bromide in the gel solution. The gel should be about 20 cm long and 10–15 cm wide and should contain one large well in the middle (7–10 cm wide) for loading the DNA sample and two small wells on each side for loading the marker. Leave an empty slot between marker and DNA sample (see Note 12).
2. Load ~100 ng of 36 kb Fosmid Control Insert DNA on each side of the gel and the crude DNA (2–5 μg) in the big well. In this way, as much as 0.5 mL can fit inside depending on the thickness of the gel.
3. Resolve the samples by gel electrophoresis: first 10–30 min at 50–60 V and then overnight for 16–18 h at 40–50 V (use 1.5 V per each cm between electrodes).



Fig. 1. Metagenomic DNA isolated from the sponge *Psammocinia* aff. *bulbosa*. *Lane 1*, Fosmid Control Insert DNA (~36 kb); *lane 2*, crude DNA (all sizes); *lane 3*, size-selected DNA (~35 kb) purified by LMP agarose electrophoresis.

4. Using a scalpel, remove a gel slice from each side of the gel containing a small part (~0.5–1 cm wide) of the well where DNA sample was loaded and the whole marker lane.
5. Stain both gel slices with ethidium bromide for 20 min with gentle shaking.
6. After 20 min, expose the stained slices to UV, and by using a clean scalpel, mark the position where the 36 kb marker is located, making sure that a bit more region is included above where usually most of the DNA of interest is found. The slice should not be wider than 5 mm. Keep in mind that only DNA from 30 to 45 kb will later be packaged.
7. Away from the UV light, realign all gel slices. According to the markings made on each side, locate and excise the DNA band of interest (at ~40 kb) in the central unstained portion of the gel. Do not expose this region to UV light.

3.2.2. Recovery of Size-Selected Metagenomic DNA from LMP

1. This step is done according to the Epicentre protocol (18) with minor modifications.
2. This procedure will recover ~25–30% of the DNA initially loaded on the LMP agarose gel.
3. Before starting, set up two water baths, one at 70°C and the other at 45°C.
4. Weigh the gel slice in a 15 mL tube. Each 1 g of solidified agarose will yield approximately 1 mL of melted agarose.

5. Melt the LMP slice by incubating the tube at 70°C for 10–15 min. Quickly transfer the tube to 45°C. Just before adding, warm the GELase 50× buffer to 45°C (this prevents the agarose from solidifying once again).
6. Add the appropriate volume of warmed GELase 50× buffer to 1× final concentration and mix gently. Carefully add 1 U (1 μL) of GELase enzyme preparation to the tube for each 300 μL of melted agarose and mix gently.
7. Quickly transfer the melted agarose solution at 45°C and incubate for at least 2 h.
8. At the end of the incubation time, transfer the reaction to 70°C for 10 min to inactivate the GELase enzyme.
9. Centrifuge at room temperature for 20–30 min to pellet any insoluble oligosaccharides (see Note 6).
10. Any "pellet" will be gelatinous, and translucent to opaque. Carefully remove the upper 90–95% of the supernatant, which contains the DNA, to a sterile 15 mL tube. Be careful to avoid the gelatinous pellet.
11. Precipitate the DNA at room temperature (RT). Add one-tenth volumes of 3 M sodium acetate (pH 7.0) and mix gently. Add 2.5 volumes of 100% ethanol equilibrated at RT. Cap the tube and mix by gentle inversion until the solution looks homogeneous. Allow precipitation for 10–15 min at RT.
12. Centrifuge the precipitated DNA for 20–30 min at RT (see Note 6). The DNA pellet has a brown-whitish appearance.
13. Carefully aspirate about 95% of supernatant from the pelleted DNA.
14. Wash the pellet 2× with ice-cold 70% ethanol and each time centrifuge for 20 min at 4°C (see Note 6). Special care must be taken not to disrupt the DNA pellet. Usually, the DNA loses its color due to coprecipitated salts in these steps.
15. After the second washing step, carefully remove the supernatant and air-dry pellet until no ethanol remains inside. This step can be shortened considerably if drying is done inside a clean bench (see Note 4).
16. Gently resuspend DNA pellet with 55 μL Tris–HCl buffer (10 mM, pH 8.5; see Notes 9 and 13).
17. Determine the DNA concentration and quality by running an aliquot (0.5–1 μL) on a 1% agarose gel using known amounts of Fosmid Control Insert DNA.

3.2.2.1. Recovery of Size-Selected Metagenomic DNA from LMP by Electroelution

As an alternative to this gel extraction protocol, one can recover the size-selected DNA by electroelution. This procedure will as well recover ~25–30% of the DNA initially loaded on the LMP agarose gel.

3.2.2.1.1. Procedure

1. Hydrate a piece of dialysis tubing (~3 cm longer than the gel slice) for ~30 min in distilled water, to remove sodium azide traces. Then, rinse the tubing inside for several times with distilled water.
2. Clip one end of the dialysis tubing and gently push gel slice inside trying to remove as much air as possible before closing the other end. Add some buffer (same as the gel, e.g., 1× TAE buffer) so that the slice is completely immersed and clip the open end. Make sure once again that there are no bubbles inside.
3. Place gel slice parallel to the electrodes and fill the electrophoresis chamber with buffer until the tubing is submerged. If some bubbles are still present, they should be positioned on top and not behind or in front of the gel slice. At the same time, make sure that the clip edges rest on the bottom.
4. Electroelute at 100 V for 2–3 h.
5. At the end, reverse the electrodes and run for 2 min at 35 V to dislodge the DNA from the dialysis tubing.
6. Open the clip, carefully remove the agarose slice (which can be restained to verify elution), and pour off the DNA-containing buffer into a Petri dish (see Note 14).
7. In a microcentrifuge, pellet (15,000 ×*g*, 30 min, RT) any agarose that you might have transferred out of the dialysis tubing.
8. DNA precipitation is then done according to the above protocol (see above DNA recovery from LMP agarose gel, starting with step 11).

Figure 1 (lane 3) shows an example of size-selected DNA recovered from LMP by electroelution.

3.2.3. Blunt-Ending of Size-Selected Metagenomic DNA

1. This step is done according to the Epicentre protocol (18) with minor changes.
2. This procedure will recover ~50–75% of the DNA initially pipetted in the blunt-ending reaction.
3. Set up the following blunt-ending reaction:

8 μL	10× end-repair buffer
8 μL	2.5 mM dNTP mix
8 μL	10 mM ATP
x μL	Size-selected metagenomic DNA ←(~ 500 ng–2 μg)
4 μL	End-repair enzyme mix (mixture of T4 DNA polymerase and T4 polynucleotide kinase)
x μL	Sterile water (add to 80 μL final reaction volume)

4. Remove all bubbles and incubate for exactly 45 min at room temperature (see Note 15).
5. Stop reaction by adding EDTA to a final concentration of 10 mM and by an additional inactivation step at 70°C for 10 min.
6. Isopropanol-precipitate blunt-ended metagenomic DNA by adding the following to the 80 μL end-repair reaction: 120 μL sterile water, 20 μL 3 M sodium acetate (pH 5.2), and 140 μL isopropanol. Mix gently by inversion. Allow precipitation to occur for 30 min at room temperature.
7. Spin in a 4°C microcentrifuge at top speed for 30 min.
8. Remove ~95% of supernatant. Be very careful during this step, since in addition to a pellet at the bottom, the sample is often distributed up on the sides of the tube. At this stage, the pellet might not be clearly visible.
9. Carefully add 500 μL of ice-cold 70% ethanol to the pellet. Try not to dislodge the pellet from the bottom of the tube.
10. Spin in a microcentrifuge at top speed for 10 min, 4°C.
11. Carefully remove all of the supernatant from the pellet by using the same level of caution as in the step above. The DNA pellet should now be visible as a thin layer at the bottom and/or on the sides of the tube.
12. Spin the sample briefly in a microcentrifuge.
13. Carefully remove any drops of liquid from the bottom of the tube.
14. Air-dry the pellet at room temperature. This step can be shortened considerably if drying is done inside a clean bench (see Note 4).
15. Add 10 μL of Tris–HCl buffer (pH 8.5) for each μg of DNA pipetted in the blunt-ending reaction.
16. Allow DNA sample to resuspend for ~30 min on ice (see Note 15).
17. Determine the concentration of the blunt-ended DNA sample by running 0.5–1 μL on an analytical 1% agarose gel with known amounts of Fosmid Control Insert DNA.

3.2.4. Ligation of Blunt-Ended Metagenomic DNA

1. This step is done according to the Epicentre protocol (*18*) with slight changes.
2. A single ligation reaction will produce ~10^3–10^6 clones depending on the quality of the insert DNA.
3. A 10:1 molar ratio of CopyControl pCC1FOS fosmid (Epicentre) to insert DNA is optimal: 0.5 μg pCC1FOS (~8.1 kb; 0.09 pmol fosmid) and 0.25 μg of DNA prepared in Subheading 3.2.3 (0.009 pmol insert DNA).

4. Combine the following reagents in the order listed and mix thoroughly after each addition:

x μL	Sterile water (add to 10 μL final reaction volume)
1 μL	10× T4 DNA ligase reaction buffer
1 μL	pCC1FOS fosmid (0.5 μg/μL)
x μL	Blunt-ended insert DNA (0.25 μg of DNA: this is optimal, more is not better)
1 μL	T4 DNA ligase

5. Incubate at 16°C overnight (16–18 h).
6. Incubate at 65°C for 10–15 min to inactivate the T4 DNA ligase.
7. Proceed immediately with the packaging reaction.

3.2.5. Packaging the CopyControl Fosmid Clones

1. This step is done according to the Epicentre protocol (18) with small changes.
2. The day before the packaging reaction, use a toothpick to inoculate 5 mL of LB broth + 10 mM $MgSO_4$ with a colony of EPI300-T1^R *E. coli* cells (Epicentre) from a fresh plate. Incubate overnight at 37°C and 200 RPM.
3. On the day of the packaging reactions, inoculate 50 mL of LB broth + 10 mM $MgSO_4$ with 5 mL of the EPI300-T1^R *E. coli* overnight culture from the previous step. Shake at 37°C, 200 RPM, to an OD_{600} = 0.8–1.0. Store the cells at 4°C until needed. They can be used directly without reheating.
4. Thaw, on ice, one tube of the MaxPlax lambda packaging extracts (Epicentre) for every 10 μL ligation reaction done (see Note 16).
5. When thawed, immediately transfer 25 μL (one-half) of each packaging extract to a second 1.5 mL microfuge tube and place on ice. Return the remaining 25 μL of the MaxPlax Packaging extract to the –80°C freezer for later use (see Note 17).
6. Add 10 μL of the ligation reaction to each 25 μL of the thawed extracts being kept on ice.
7. Mix by pipetting the solutions several times. Avoid the introduction of air bubbles. Briefly centrifuge the tubes to get all liquid to the bottom.
8. Incubate the packaging reaction at 30°C for 90 min. Use a water bath to heat the reaction because heat transfer in incubators that do not have contact with the reaction (air) is too slow.

9. After the 90 min packaging reaction is complete, add the remaining 25 μL of MaxPlax lambda packaging extract from step 5 to the tube.
10. Incubate the reaction for an additional 90 min at 30°C.
11. At the end of the second 90 min incubation, add PDM to 1 mL final volume and mix gently. The packaging reaction can be stored at 4°C until needed (in any case not longer than 1 day). Proceed with the titering of the phage particles (packaged fosmid clones) and then plate the fosmid library.

3.2.6. Titering the Packaged CopyControl Fosmid Clones

1. Before plating the library, it is highly recommended that the titer of the phage particles (packaged CopyControl fosmid clones) is determined. This will aid in determining the number of clones inside the library.
2. The number of fosmid clones required to reasonably ensure that any given DNA sequence is contained within the library can be determined using the formula $N=\ln(1-P)/\ln(1-f)$. P is the desired probability (expressed as a fraction); f is the proportion of the genome contained in a single clone; and N is the required number of fosmid clones. For example, the number of clones required to ensure a 99% probability of a given DNA sequence of *E. coli* (genome = 4.7 Mb) being contained within a fosmid library composed of 40 kb inserts is: $N=\ln(1-0.99)/\ln[1-(4\times10^4 \text{ bases}/4.7\times10^6 \text{ bases})]=-4.61/-0.01=461$ clones.
3. Make serial dilutions of the 1 mL packaged phage particles (see previous section) into PDB as follows:
   (a) 1:10 Dilute 10 μL of packaged phage into 90 μL of PDB.
   (b) 1:10^2 Dilute 10 μL of the 1:10 dilution into 90 μL of PDB.
   (c) 1:10^3 Dilute 10 μL of the 1:10^2 dilution into 90 μL of PDB.
4. Add 10 μL of each above dilution, individually, to 100 μL of EPI300-T1^R *E. coli* host cells prepared in the step 3 of the previous section. Incubate for 20 min at 37°C in a water bath. Store the rest of the cells at 4°C until 2nd day, as they will be used again when plating the entire library.
5. Spread the infected EPI300-T1^R *E. coli* cells on LB plates containing 12.5 μg/mL chloramphenicol (see Note 18).
6. Count colonies and calculate the titer of the packaged phage particles with the formula: (number of colonies) × (dilution factor) × 100.

3.3. Metagenomic DNA Library Plating and Screening by PCR

1. This section is based on the protocol provided in the paper by Hrvatin and Piel (17). Due to several advantages mentioned in the introduction of the methods section, the library is not plated but rather distributed in 2 mL screw-cap vials.

2. Autoclave as many 2 mL screw-cap vials as you will need. You will need one tube for each 1,000 colony forming units (cfus). Autoclave the tubes and the caps separately.
3. Make 1× LB broth with 5 g/L of SeaPrep agarose (see Notes 3 and 19). One liter of this medium can be used to grow ~10^6 colonies.
4. Mix the 1 mL packaging reaction from the previous step with 10 mL of EPI300-T1^R *E. coli* host cells prepared 1 day before (see above). Incubate for 20 min at 37°C in a water bath.
5. Based on the titer of the phage particles determined above, mix the 11 mL reaction from the previous step with the corresponding volume of LB containing 5 g/L of SeaPrep agarose, to obtain ~1,000 cfus/mL of medium. Add the appropriate antibiotics (12.5 μg/mL chloramphenicol) and then mix well using a magnetic stirrer. Avoid foaming.
6. In the meantime, set up in a fixed stand placed on wet ice as many 2-mL screw-cap vials as needed. The vials should have no cap and the level of the ice should be above half of the tube.
7. Distribute 1 mL aliquots of the mixture from step 4 to obtain in each tube a density of ~1,000 cfus. Each of them constitutes a library pool.
8. Keep the vials on wet ice for 1 h. The ice level should be higher than the level of the medium in the tube. Put the caps on the vials, taking care not to shake them too much (see Note 3).
9. Gently remove the vials from the wet ice and place them at 37°C for 16–18 h. Figure 2 shows a 400,000 member metagenomic DNA library distributed in screw-cap vials with semiliquid medium.



Fig. 2. Metagenomic DNA library consisting of ~400,000 cfus distributed in screw cap vials with semiliquid medium. Clones are visible as white dots growing spatially.

10. Colonies should be visible inside as white dots growing spatially (Fig. 2). Ideally, there should be no precipitate inside the vials (see Note 3).
11. Vortex each vial for several seconds until the suspension becomes homogeneous.
12. Take, if necessary, an aliquot for PCR (~20 μL) and small aliquots for superpools (see below, library screening section).
13. Add 0.5 mL of 50% glycerol and mix by inversion.
14. The cultures can then be safely frozen at −80°C.
15. A large metagenomic library can contain up to several million clones. To avoid screening of thousands of individual pools, PCRs with combined aliquots from several pools (termed superpools) still give clear results in spite of the low copy number of pCC1FOS. The following steps are performed by using standard cryoboxes with nine rows and nine columns.
16. Prepare column and row superpools by mixing 30 μL of each pool from a column or a row, respectively. Analyze these superpools (each containing ~ 9,000 clones) in a first round of PCR. For a 25 μL PCR setup, 0.5 μL of bacterial suspension is used directly and lysed in an initial 5 min 95°C step.
17. In case a superpool yields the expected amplicon, after another round of PCR with the nine individual pools constituting the superpool, distribute the positive mixture among 30 vials with 1 mL of fresh gel medium at a concentration of ~ 100 cfus. In doing such dilutions, take into account that 1 mL of semiliquid culture grown overnight at 37°C will have an OD_{600} of ~0.1, which corresponds to $\sim 2 \times 10^8$ cfus. Thus, for obtaining a subpool with a density of ~100 cfus/mL/tube, dilute the pool 10^4-fold and take from the dilution a volume of 150 μL, mix it with 30 mL of semiliquid medium (containing 12.5 μg/mL chloramphenicol) and distribute the mixture to 30 tubes (see Note 20).
18. Incubate mixtures at 37°C overnight, and screen them by PCR. Dilute and distribute the positive subpool among 20 vials at a density of ~15 cfus in a similar manner as described above. In this case, dilute each subpool 10^5-fold, take 150 μL and mix with 20 mL SeaPrep medium with 12.5 μg/mL chloramphenicol (see Note 20).
19. In the final round of screening, plate an aliquot of the positive pool on regular agar plates containing 12.5 μg/mL chloramphenicol to generate individual colonies, which are then analyzed by colony PCR. In this way, a positive clone can be isolated from a library of ~400,000 clones in a bit less than 5 days (17). A general scheme for screening highly complex libraries with this procedure is provided in Fig. 3.

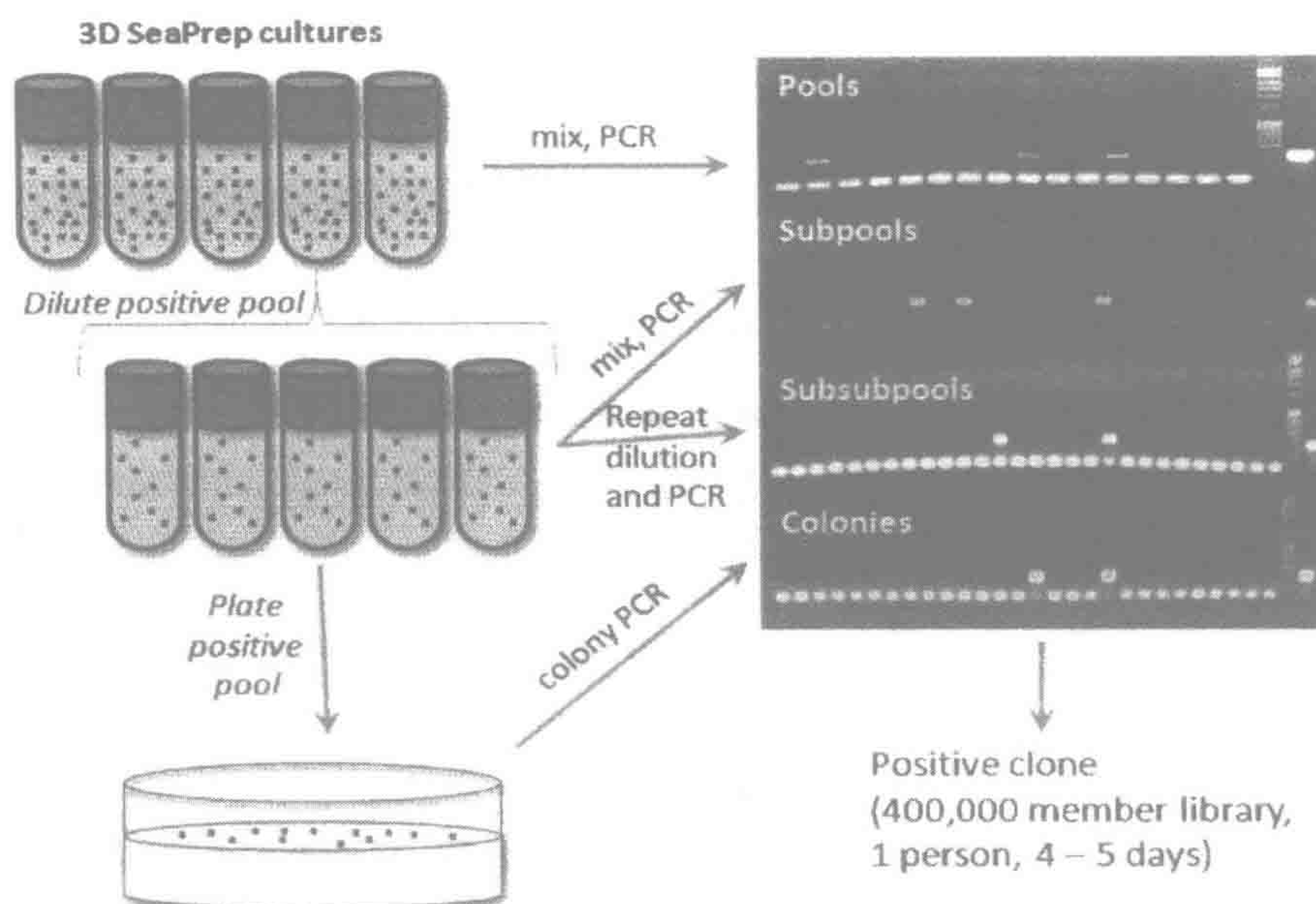


Fig. 3. Isolation of rare clones from highly complex libraries. Pools of up to 1,000 clones are grown in 2-mL screw-cap vials containing a semiliquid medium. Superpools (mixed pools) containing ~9,000 clones are PCR-screened. A second round of PCR is performed on the pools from which the positive superpool is derived. The positive pools are then further diluted to subpools consisting of ~100 cfus/tube and subsequently to ~15 cfus/tube. The positive clone is at the end isolated from the plate by colony PCR.

4. Notes

1. Isolation protocol works fine with samples stored either at 4°C or at –80°C.
2. Because CTAB crystals form at room temperature, the solution has to be warmed at 60°C prior use.
3. In case colonies do not grow spatially inside the semiliquid cultures and precipitate at the bottom of the tube, the concentration of the SeaPrep agarose can be increased from 0.5 to 0.6%. However, too much shaking during cap screwing might still result in a precipitate. In this way, fast-growing colonies attached to the bottom of the tube can tend to overgrow others, and this might lead to unsuccessful screenings.
4. If the DNA yield is low or if there is no DNA at all, it could be worthwhile to test whether omitting the CTAB improves the situation.
5. If the DNA yield is rather low, the lysis step can be extended to 30 min.

6. Unless otherwise stated, all centrifugation steps are performed at 10,000 × g, 4°C, for 5 min.
7. Be very careful not to overdry genomic DNA pellets, since this makes solubilization very difficult.
8. A sample of the crystallized fraction can also be run on the gel to check for DNA loss.
9. If for some reason, the DNA pellet does not dissolve easily despite all this, the sample can be incubated shortly (10–15 min) at 50°C.
10. An uneven migration of DNA upon electrophoresis indicates the presence of polysaccharides.
11. Use TAE buffer instead of TBE as borate ions can detrimentally affect the activity of many enzymes, such as GELase, which is used in the subsequent DNA gel extraction step.
12. LMP agarose gels are very hard to handle, and they break very easily. It takes at least 1 h for the gel to solidify at room temperature.
13. Do not use Tris–EDTA buffer, since the EDTA may inhibit or lower efficiency of subsequent enzymatic reactions (e.g., end repair, ligation).
14. If low yields are suspected, add another 200 μL of 1× TAE buffer to the empty tubing and rinse several times. Final volume should be as low as possible.
15. Avoid longer incubation times at higher temperatures, as T4 DNA polymerase besides its template-directed 5′→3′ DNA polymerase activity exhibits also a potent 3′→5′ exonuclease activity.
16. It takes ~30 min for the extracts to thaw on ice.
17. Do not expose the MaxPlax Packaging Extracts to dry ice or other CO_2 source.
18. The pCC1FOS fosmid contains a chloramphenicol selectable marker. Incubate at 37°C overnight to select for the CopyControl fosmid clones.
19. Make sure the LB is stirred vigorously when the SeaPrep agarose is added otherwise large clumps will form. Try dissolving the agarose by heating in the microwave until no larger clumps are present. The rest of the dissolution will be done by the autoclave. Autoclave at 121°C for 20 min.
20. The calculation applies only for a total dilution volume of 1 mL. All dilutions are done with LB broth medium. Estimate titer of each dilution (e.g. subpool) by plating 100 μL of mixture at the beginning, in the middle, and at end of the distribution of 1 mL aliquots to the screw-cap vials. The plates should contain 12.5 μg/mL chloramphenicol.

Acknowledgements

The authors would like to thank the current and former laboratory members Daniel Butzke, Sinişa Hrvatin, Katja Fisch, Nina Heycke, Dequan Hui, and Sonia van der Sar for contributing to the development of these methods, Sean Brady for advice on metagenomic techniques, and Shigeki Matsunaga, Nobuhiro Fusetani, Phil Crews, and Vicky Webb for sponge samples. The work was supported by grants from the DFG, BMBF, JSPS, and NSF, and scholarships from the Alexander von Humboldt Foundation, and the DAAD.

References

1. "Invertebrate." Student Encyclopaedia. Encyclopaedia Britannica (http://kids.britannica.com/comptons/article-202857/invertebrate).
2. Ogle, W. (1882) Aristotle. On the parts of animals. Kegan Paul, Trench & Co., London.
3. Grant, R.E. (1836) The Encyclopedia of Anatomy and Physiology. Sherwood, Gilbert and Piper, London.
4. Rigby, J.K. (1983) Introduction to the Porifera. In: Rigby, J.K. and Steam, C.W. Ed. Sponges and Spongiomorphs. Notes for a Short Course, Vol. 7. Stud. Geol. Univ., Tennessee, pp. 1–11.
5. Love, G.D., Grosjean, E., Stalvies, C., Fike, D.A., Grotzinger, J.P., Bradley, A.S., et al. (2009) Fossil steroids record the appearance of Demospongiae during the Cryogenian period. *Nature* 457, 718–721.
6. Blunt, J.W., Copp, B.R., Hu, W.P., Munro, M.H., Northcote, P.T., and Prinsep, M.R. (2009) Marine natural products. *Nat Prod Rep* 26, 170–244.
7. Piel, J. (2004) Metabolites from symbiotic bacteria. *Nat Prod Rep* 21, 519–538.
8. Donadio, S., Monciardini, P., and Sosio, M. (2007) Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat Prod Rep* 24, 1073–1109.
9. Unson, M.D. and Faulkner, D.J. (1993) Cyanobacterial symbiont biosynthesis of chlorinated metabolites from *Dysidea herbacea* (Porifera). *Experientia* 49, 349–353.
10. Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N., et al. (2004) Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc Natl Acad Sci USA* 101, 16222–16227.
11. Munro, M.H., Blunt, J.W., Dumdei, E.J., Hickford, S.J., Lill, R.E., Li, S., et al. (1999) The discovery and development of marine compounds with pharmaceutical potential. *J Biotechnol* 70, 15–25.
12. Salomon, C.E., Magarvey, N.A., and Sherman, D.H. (2004) Merging the potential of microbial genetics with biological and chemical diversity: an even brighter future for marine natural product drug discovery. *Nat Prod Rep* 1, 105–121.
13. Webster, N.S., Wilson, K.J., Blackall, L.L., and Hill, R.T. (2001) Phylogenetic diversity of bacteria associated with the marine sponge *Rhopaloeides odorabile*. *Appl Environ Microbiol* 67, 434–444.
14. Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M. (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5, R245–R249.
15. Piel, J. (2006) Bacterial symbionts: prospects for the sustainable production of invertebrate-derived pharmaceuticals. *Curr Med Chem* 13, 39–50.
16. Rogers, S.O. and Bendich, A.J. (1985) Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Mol Biol* 5, 69–76.
17. Hrvatin, S. and Piel, J. (2007) Rapid isolation of rare clones from highly complex DNA libraries by PCR analysis of liquid gel pools. *J Microbiol Methods* 68, 434–436.
18. EPICENTRE Biotechnologies. CopyControl™ Fosmid Library Production Protocol (http://www.epibio.com/pdftechlit/171Pl079.pdf).

Chapter 18

Screening for Novel Antibiotic Resistance Genes

Mindy G. Brown, Elizabeth H. Mitchell, and David L. Balkwill

Abstract

Knowledge of novel antibiotic resistance genes aids in the understanding of how antibiotics function and how bacteria fight them. This knowledge also allows future generations of an antibiotic or antibiotic group to be altered to allow the greatest efficacy.

The method described here is very simple in theory. The bacterial strains are screened for antibiotic resistance. Cultures of the strain are grown, and DNA is extracted. A partial digest of the extraction is cloned into *Escherichia coli*, and the transformants are plated on selective media. Any colony that grows will possess the antibiotic resistance gene and can be further examined. In actual practice, however, this technique can be complicated. The detailed protocol will need to be optimized for each bacterial strain, vector, and cell line chosen.

Key words: Antibiotic resistance, Shotgun cloning, Sequencing

1. Introduction

In nature, bacteria are thought to produce antibiotics to attack neighboring bacteria as a method to preserve their resources (1). When resources are limited, a bacterial colony will produce an antibiotic to destroy or inhibit neighboring bacteria, thereby limiting competition for the scarce resources. In order for this strategy to be effective, the bacteria producing the antibiotic must be able to survive and do so by possessing mechanisms of resistance to the antibiotic they produce. These mechanisms can be transferred to other bacteria, and this has led to the clinical problem we now face, where antibiotic resistance in pathogenic strains makes it difficult to fight bacterial infections (2). This has led to an increasing threat to global public health by confounding treatment of infections caused by virtually all major pathogens (2–4).

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668,
DOI 10.1007/978-1-60761-823-2_18, © Springer Science+Business Media, LLC 2010

The resistance problem is compounded by the administration and misuse of antibiotics. Physicians often prescribe antibiotics when they are not needed or without knowing the cause of their patients' illnesses. Patients will also misuse an antibiotic by not completing the entire course of treatment, allowing the bacteria that are resistant to low doses of the antibiotic to survive. In all cases, the surviving bacteria can now proliferate and transfer their resistance mechanisms to other pathogenic bacteria. Agricultural use of antibiotics is also causing a rise in antibiotic resistance. Livestock feed is regularly treated with antibiotics as a preventative measure rather than as a treatment for illness. One study estimated that 8.5 million kilogram of antibiotics is used annually in the United States (5). Much of the antibiotics used in this way will ultimately end up in receiving water via manure as fertilizer for crops (5), or in meat and milk through animal products (6), causing unnecessary exposure to humans.

There are two different ways that bacteria develop resistance to antibiotics: through mutations in certain functional genes or by acquiring new genes that provide a resistance mechanism. The type of resistance that will be examined in this chapter is the possession of a resistance gene. The method utilized to study resistance genes is a variation of shotgun cloning. DNA extractions from an antibiotic-resistant isolate are partially digested, cloned into an expression vector, and transformed into *Escherichia coli*. The transformed cells are plated on media containing the antibiotic to which the isolate is resistant. Any colony that grows on these plates contains the gene for resistance. The gene can then be further studied from the clone. One of the major limitations to this method is the inability to test all antibiotics. Some antibiotics, such as vancomycin, are ineffective against Gram-negative bacteria because they cannot penetrate the outer cell wall. *E. coli* cells are the only commonly available competent cells, thereby preventing any screening for resistance to antibiotics that are only effective against Gram-positive bacteria.

2. Materials

2.1. DNA Extraction

1. RNase A: 100 mg/mL dissolved in H_2O and should be made fresh.
2. Buffer B1 (Bacterial lysis buffer): 50 mM Tris–HCl pH 8.0, 50 mM EDTA pH 8.0, 0.5% Tween 20, 0.5% Triton X-100. Store at 2–8°C; equilibrate to room temperature before use.
3. Lysozyme stock: 100 mg/mL dissolved in H_2O; store at –20°C for up to 1 month.
4. Proteinase K: 20 mg/mL and can be ordered from Qiagen.

5. Buffer B2 (Bacterial lysis buffer): 3 M guanidine HCl, 20% Tween 20; store at 2–8°C; equilibrate to room temperature before use.
6. Lysis Matrix A tubes from MP Biomedical.
7. FastPrep-24 instrument (MP Biomedicals, Solon, OH).

2.2. DNA Preparation

1. One blunt-end enzyme, e.g., HaeIII, HpyCH4V, RsaI, AluI, DpnI CviKI-1, EcoRV, or BsaAI (New England Biolabs Inc., Ipswitch, MA).
2. 5 M NaCl.

2.3. Cloning of Antibiotic Resistance Determinant

1. pEZseq vector: purchased from Lucigen Corporation, Middleton, WI. Another suitable vector with a promoter can be used.
2. Competent cells – the following procedure uses the cells that are provided with the kit: 10G Elite from Lucigen. Other competent cells can be used.
3. Bio-Rad Micropulser.
4. YT agar plates (per liter): 8 g Bacto-tryptone, 5 g yeast extract, 5 g NaCl, 15 g agar. Add X-gal to a final concentration of 50 μg/mL. Add IPTG to the concentration required for the cell line.

2.4. DNA Sequencing

1. QIAprep Spin Miniprep Kit (Qiagen).
2. TB medium with 30 μg/mL kanamycin (per liter): 11.8 g bacto-tryptone, 23.6 g yeast extract, 9.4 g dipotassium hydrogen phosphate (anhydrous), 2.2 g potassium dihydrogen phosphate (anhydrous), 0.4% glycerol. Mix all ingredients except glycerol. After autoclaving and cooling, add 8 mL filter-sterilized 50% glycerol per liter prior to use. When medium is cool, add antibiotic.
3. Restriction enzymes: AlwnI and DraI (NEB).
4. 454 Life Sciences sequencer and Newbler sequences assembly software (454 Life Sciences).

3. Methods

The following method utilizes selective media to allow only colonies with a specific property (antibiotic resistance) to survive. The method has five components: DNA extraction, partial digest of DNA, cloning of the partial digest, colony sequencing, and computational analysis of DNA sequences.

This method can be adopted for any gene that allows the bacterial colony to survive on selective media.

3.1. Total DNA Extraction

1. Strains are grown in a suitable liquid medium and collected by centrifugation during early log-phase growth.
2. Perform a total DNA extraction. The method outlined below is a modification of the Genomic Tip Kit (Qiagen, Valencia, CA) in conjunction with the FastPrep-24 instrument (see Note 1).
   (a) For each sample, add 22 μL of RNase A solution to 11 mL of Buffer B1.
   (b) Resuspend bacterial pellet in 11 mL aliquot from step a by vortexing.
   (c) Add 300 μL of lysozyme stock solution and 500 μL of Proteinase K stock solution. Incubate at 37°C for at least 30 min (see Note 2).
   (d) Add 4 mL of Buffer B2 and mix. Incubate at 50°C for 30 min.
   (e) Add samples to Lysis Matrix A tubes. Several tubes will be needed to aliquot all the sample.
   (f) Place tubes in the FastPrep-24 instrument and lyse at 4.0 min/s for 60 s.
   (g) Centrifuge samples at 12,000 × *g* for 1 min and then recombine the supernatant for each sample in 15-mL tubes.
   (h) Precipitate DNA by adding 1.4 mL isopropyl alcohol and centrifuging at 5,000 × *g* for 15 min at 4°C.
   (i) Remove the supernatant and wash pellets with 1 mL of cold 70% EtOH and recentrifuge at 5,000 × *g* for 10 min at 4°C.
   (j) Aspirate EtOH and dry pellet for 5–10 min (until translucent) and resuspend in 0.1–2 mL of 1× TE.

3.2. DNA Preparation

1. DNA extractions from each strain are partially digested with one of the following blunt-end enzymes (see Notes 3 and 4):
   (a) HaeIII, HpyCH4V, RsaI, AluI, DpnI CviKI-1, EcoRV, or BsaAI.
   (b) The digest reaction should include approximately 1,200 ng of DNA.
   (c) The reaction time and the enzyme concentration will need to be optimized for each strain.
   (d) Aliquots of the digests are analyzed on a 0.8% agarose gel, and the digest with the best shearing pattern for each strain is selected (see Note 5).

2. The DNA from the selected digest is purified and concentrated with an ethanol precipitation.
   (a) Add 1/10th volume 5 M NaCl.
   (b) Add 1 volume isopropanol or 2 volumes ethanol.
   (c) Mix well and place at −20°C for 30 min.
   (d) Let warm to 0°C.
   (e) Centrifuge 15 min at a speed fast enough to form pellet and remove supernatant.
   (f) Gently rinse the pellet with 70% EtOH and recentrifuge. Repeat once.
   (g) Aspirate liquid and air-dry pellet until pellet is dry.
   (h) Resuspend pellet to the required concentration in a suitable buffer.

3.3. Cloning of Antibiotic Resistance Determinant

1. Ligate purified digests (see Subheading 3.2, step 2) into the pEZSeq vector (see Notes 4, 6, and 7).
   (a) The ligation reaction is performed using 250–500 ng of the DNA insert according to the manufacturer's specifications.
2. The vector is now transformed into electrocompetent cells (see Note 8).
   (a) The cells used are provided with the 10G Elite kit and are transformed according to the kit instructions using a Micropulser.
3. Recovered cells are plated onto selective YT agar plates.
4. Controls should be performed as stated in the protocol book provided with the kit.

3.4. DNA Sequencing

1. Clones that grow on the selective YT agar plates are the colonies that possess a vector that has picked up a resistance gene. These colonies are selected and grown overnight at 37°C in 10 mL liquid TB medium containing 30 μg/mL kanamycin.
   (a) Five milliliter of each culture is centrifuged at 2,500 × g for 10 min, and the pellet can be used for plasmid DNA extraction following the protocol in the QIAprep Spin Miniprep Kit.
   (b) The other 5 mL of the culture is centrifuged at 200 × g for 10 min. Resuspend the pellet in TB medium with 20% glycerol; then store at −80°C.
2. Cloned plasmid DNA is now screened using a double digest with enzymes AlwnI and DraI specific to the vector used.
   (a) The digest is run on a 1% agarose gel to identify clones containing an insert.

3. The DNA is now ready for sequencing. Several methods can be used, and in our analysis of a novel tetracycline resistance gene (7), we chose to use the 454 Life Sciences sequencer because some of the inserts were quite large.
   (a) The Sequencing was performed as described in Margulies et al. (8) at the University of Florida (UF) Interdisciplinary Center for Biotechnology Research (ICBR).
   (b) After sequencing, the UF ICBR assembled the sequences using Newbler sequences assembly software.

3.5. Computational Analysis

1. The sequences first need to be evaluated for homology to known genes through the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST).
2. The DNA sequences will need to be converted to protein (amino acid) sequences and aligned to other genes to determine relatedness.

4. Notes

1. This method was designed and optimized for bacterial strains isolated from the deep terrestrial subsurface. Some environmental strains require modifications in the DNA extraction protocol to disrupt thick cell walls. If a suitable DNA extraction method is already in use, skip to Subheading 3.2.
2. Depending on the bacterial strain used, you may need to let this incubation go longer. The sample should be very homogeneous. An incubation period of 1.5 h is not uncommon.
3. The easiest method of setting up a partial digest is to set up a dilution series of the enzyme.
4. A nice feature of this method is that it can be adopted for different vectors and restriction enzymes. A blunt-end vector was chosen for this protocol for ease of choosing restriction enzymes.
5. Digests for tetracycline-resistant bacterial strains were selected if the shearing pattern on an agarose gel was between 4 and 10 kb. This size was selected because the genes that provide resistance to tetracycline are in this size range. A larger or smaller band size can be used, but the band does need to show significant shearing.
6. When selecting a vector, it is important to screen the vector for antibiotic resistance to your antibiotic of choice. Many vectors will have resistance to antibiotics that are not listed.

7. It is important that the vector have a promoter to provide transcription of the gene of interest. With this approach, the insert need not possess its own promoter, which could be much further upstream from the region that is inserted.
8. There are several aspects of the electroporation step that can provide an error. The first is the micropulser. The leads and contacts on the equipment cast loose very easy and should be checked before each use. Check to make sure that the cuvette is properly seated in the equipment as well. The cuvettes that are recommended for the unit are also the ones that work the best. The second aspect of this step that is critical is the salt concentration of the sample. If the salt concentration is too high, the electroporation will not work; purifying the partial digest ensures minimal salt concentration.

References

1. Seveno, N.A., Kallifidas, D., Smalla, K., Dirk van Elsas, J., Collard, J.-M., Karagouni, A.D., et al. (2002) Occurrence and reservoirs of antibiotic resistance genes in the environment. *Rev Med Microbiol* 13, 15–27.
2. Wise, R. (2004) The 2003 garrod lecture: The relentless rise of resistance? *J Antimicrob Chemother* 54, 306–310.
3. Levy, S.B. (2002) The 2000 Garrod lecture. Factors impacting on the problem of antibiotic resistance. *J Antimicrob Chemother* 49, 25–30.
4. Levy, S.B. and Marshall, B. (2004) Antibacterial resistance worldwide: causes, challenges and responses. *Nat Med* 10, S122–S129.
5. Boxall, A.B., Kolpin, D.W., Halling-Sorensen, B., and Tolls, J. (2003) Are veterinary medicines causing environmental risks? *Environ Sci Technol* 37, 286A–294A.
6. Khachatourians, G.G. (1998) Agricultural use of antibiotics and the evolution and transfer of antibiotic-resistant bacteria. *CMAJ* 159, 1129–1136.
7. Brown, M.G., Mitchell, E.H., and Balkwill, D.L. (2008) Tet 42, a novel tetracycline resistance determinant isolated from deep terrestrial subsurface bacteria. *Antimicrob Agents Chemother* 52, 4518–4521.
8. Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380.

Chapter 19

Novel Metal Resistance Genes from Microorganisms: A Functional Metagenomic Approach

José E. González-Pastor and Salvador Mirete

Abstract

Most of the known metal resistance mechanisms are based on studies of cultured microorganisms, and the abundant uncultured fraction could be an important source of genes responsible for uncharacterized resistance mechanisms. A functional metagenomic approach was selected to recover metal resistance genes from the rhizosphere microbial community of an acid-mine drainage (AMD)-adapted plant, *Erica andevalensis*, from Rio Tinto, Spain. A total of 13 nickel resistant clones were isolated and analyzed, encoding hypothetical or conserved hypothetical proteins of uncertain functions, or well-characterized proteins, but not previously reported to be related to nickel resistance. The resistance clones were classified into two groups according to their nickel accumulation properties: those preventing or those favoring metal accumulation. Two clones encoding putative ABC transporter components and a serine O-acetyltransferase were found as representatives of each group, respectively.

Key words: Metal resistance, Nickel, Microorganisms, Bacteria, Metagenomic, Acid-mine drainage, Rhizosphere, Rio Tinto

1. Introduction

Toxic metals and metalloids are broadly distributed in our planet, and microbes have been exposed and adapted to them since the beginning of life. A number of metals are needed as micronutrients and play essential biochemical roles such as catalysts, enzyme cofactors, activity in redox processes, and stabilizing protein structures (1). However, most of the metals become toxic when they accumulate above normal physiological concentrations by the action of unspecific and constitutively expressed transport systems. Intracellular metals can disrupt membrane integrity, inhibit transport systems, block functional groups of enzymes,

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668, DOI 10.1007/978-1-60761-823-2_19, © Springer Science+Business Media, LLC 2010

and displace essential metals from their natural binding sites. Thus, microorganisms have developed mechanisms to be protected from high metal concentrations, and to regulate their intracellular concentration. Six metal resistance mechanisms have been described: exclusion by permeability barrier, intra- and extracellular sequestration, active transport efflux pumps, enzymatic detoxification, and reduction in the sensitivity of cellular targets to metal ions (2, 3)

Enzymes involved in metal resistance have a potential use in bioremediation, and therefore, they have significant biotechnological and environmental importance. Our knowledge about metal resistance mechanisms is based on cultured microorganisms, and most of the culture-independent methods to rescue genes from environmental samples, including uncultured microorganisms, are based on the amplification or detection of sequences similar to those of previously known genes (4–6). Nevertheless, this approach is not useful for finding novel mechanisms of metal resistance, which are still undiscovered in nature, and a different and nonbias approach should be used. We describe a culture-independent method based on the functional analysis of libraries containing environmental DNA (metagenome) from a metal-enriched environment, the Rio Tinto, an AMD environment enriched in heavy metals. The microbial DNA from the rhizosphere was used to construct small insert metagenomic libraries by using a direct lysis method for the extraction of DNA to avoid losing novel functional diversity. These libraries were screened for nickel resistance, and 13 different clones carrying resistance determinants have been identified, some of them similar to previously identified genes but others with no matches in known databases. This approach has proved to be useful to identify novel genes involved in metal (nickel) resistance, some of them not previously described (7).

2. Materials

2.1. Bacterial Strains and Media

1. Electrocompetent DH10B Cells (Invitrogen, Paisley, UK).
2. Chemically competent *Escherichia coli* DH10B cells are made and transformed by a calcium chloride protocol (8).
3. Luria–Bertani (LB) agar and broth medium supplemented with the appropriate antibiotic and/or metal (Laboratorios Conda, Torrejón de Ardoz, Spain).
4. pBluescript SKII$^+$ plasmid (Stratagene, La Jolla, CA, USA).
5. Ampicillin and kanamycin (Sigma, Steinheim, Germany) are dissolved at 50 mg/mL in water and stored in single-use aliquots at –20°C.

6. 90-mm Petri plates.
7. Sterile inoculation spreader (Sarstedt, Nümbrecht, Germany).
8. SOC medium (Invitrogen).
9. 40% Glycerol (w/v) (Merck, Whitehouse Station, NJ, USA). It is diluted to a final concentration of 10% (w/v) with LB to store cells at –80°C.
10. $NiSO_4$ (Sigma) is dissolved in sterile water at 1 M and stored at room temperature.

2.2. Molecular Procedures

1. FastPrep Instrument and BIO101 FastDNA Spin kit for soil (Qbiogene, Carlsbad, CA, USA) to purify genomic DNA from soil samples.
2. Low melting temperature SeaPlaque agarose (Cambrex Bio Science Rockland, Inc., Rockland, ME, USA).
3. Agarose D-1 LE (low EEO) powder (Laboratorios Conda, Torrejón de Ardoz, Spain).
4. 1× TAE and 1× TBE electrophoresis buffers from Invitrogen.
5. Horizontal electrophoresis apparatus including power supply, gel plates, and combs (Biorad, Hercules, CA, USA).
6. EDTA 0.5 M (Sigma).
7. 10% SDS (Gibco, Paisley, UK).
8. Proteinase K from Invitrogen.
9. Phenol (Sigma).
10. Phenol:chloroform:isoamyl alcohol (25:24:1) (Sigma).
11. 5 M sodium chloride (0.2 M final concentration) (Fluka, Buchs, Switzerland).
12. 100% ethanol (Merck).
13. 3 M sodium acetate pH 5.2.
14. Qiaquick Extration Gel (Qiagen, Hilden, Germany).
15. Sterile surgical blades.
16. Qiaprep Spin Miniprep kit (Qiagen).
17. 96-well flat-bottom block (Qiagen).
18. Ethidium bromide solution at 10 mg/mL from Invitrogen.
19. 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) at 40 mg/mL (in dimethylformide) (Duchefa Biochemie, Haarlem, The Netherlands).
20. Isopropyl-beta-D-thiogalactopyranoside (IPTG) at 100 mM (filtre sterilized in H_2O) from Roche (Mannheim, Germany).
21. DNA Ladder with a separation range between 200 and 10,000 bp, Hyperladder I (Bioline, London, UK).
22. Low-temperature water bath or heating block.

23. Restriction endonucleases *Sau3*AI (4 U/μL), *Bam*HI (10 U/μL), *Xba*I (10 U/μL), and *Eco*RI (10 U/μL) (provided from Roche) plus appropriate buffer.
24. Calf intestinal alkaline phosphatase 1 U/μL (Invitrogen) plus appropriate buffer.
25. T4 DNA ligase 1 U/μL (Roche) plus appropriate buffer.
26. *Pfu* Turbo DNA polymerase (Stratagene).
27. GPS-LS Genome Priming System (New England Biolabs, Ipswich, MA, USA).
28. Micropulser (Bio-Rad) electroporation apparatus and 1-mm electroporation cuvettes.
29. ABI PRISM dye terminator cycle-sequencing ready-reaction kit (Perkin-Elmer) and an ABI PRISM 377 sequencer (Perkin-Elmer).

3. Methods

3.1. DNA Isolation

1. 450 mg of rhizosphere and soil adhered to the roots are processed with the FastPrep Instrument and the BIO101 FastDNA Spin kit for soil, according to the manufacturer's recommendations with no further treatment (see Note 1). To increase the efficient elution of the DNA, the elution step is done twice with DES (DNase/Pyrogen Free Water), the first with 100 μL and the second with 50 μL. Approximately 33.3 μg of DNA per gram of rhizosphere sample was obtained.
2. The quality and integrity of the DNA are checked in agarose gels (0.8%) and measured by absorbance at 260, 230, and 280 nm (Abs 260, 230, and 280) (see Note 2).

3.2. DNA Fragmentation

1. To establish the conditions for restriction endonuclease partial digestion of metagenomic DNA, prepare a 100 μL reaction mixture optimal for the restriction enzyme (*Sau3*AI), containing 1 μg of metagenomic DNA. Make up the reaction mixture by adding the reagents into a 1.5-mL sterile microcentrifuge tube: 10 μL of 10× restriction endonuclease buffer, × μL (1 μg) of metagenomic DNA and add water up to a total volume of 99 μL.
2. Dispense 20 μL of the reaction mixture into microfuge tube 1 and 10 μL into the remaining tubes 2–9. Place all tubes on ice.
3. Add 1 μL (equivalent to a known amount of enzyme units, e.g., 1 U) of restriction endonuclease to microfuge tube 1, mix gently by pipetting up and down and place on ice.

4. Pipette 10 μL from tube 1 into tube 2 by using a fresh tip, mix gently, and place on ice.
5. Add 1 μL of the restriction endonuclease and mix by pipetting up and down. To dilute the enzyme by 50% between each tube, similar serial dilutions from tube 2 to 9 are carried out. Keep all the tubes on ice.
6. Incubate at 37°C for 1 h in a heat block before keeping the tubes on ice.
7. To inactivate *Sau3*AI, incubate at 65°C for 20 min in a heating block. Let the tubes cool at room temperature.
8. Analyze the efficacy of the digestion by gel electrophoresis along with an appropriate DNA molecular weight marker (e.g., Hyperladder I) on a low melting temperature agarose gel run in 1× TAE at 4 V/cm for 30 min.
9. Compare the pattern of digestion in each sample to the DNA molecular weight marker to determine which amount of restriction endonuclease produced more quantity of DNA fragments of the desired length.
10. To scale up the reaction mixture with more metagenomic DNA or number of samples, the selected amount of restriction enzyme is obtained with serial dilutions in its provided buffer. 1 μL from the correct dilution is added to the reaction mixture as outlined in step 3, and the protocol is followed as described above.
11. Excise the range of DNA fragments of appropriate length (from 1 to 8 kb) directly from the agarose gel with a sterile surgical blade and collect them into a 1.5-mL microcentrifuge tube.
12. Purify the DNA from the agarose gel with the Qiaquick Extration Gel. Add the recommended 0.5 mL of Buffer QG to eliminate all traces of agarose. Let the column stand for 5 min after addition of Buffer PE. In the elution step, add 50 μL of Buffer EB to the center of the membrane and let the column stand for 10 min before centrifugation. The resulting solution is measured by Abs 260, 230, and 280.

3.3. Vector Preparation

1. Extract the pBluescript SKII$^+$ vector with the Qiaprep Spin Miniprep kit from an overnight culture of DH10B.
2. Digest 10 μg of the plasmid with 50 U of *Bam*HI overnight at 37°C in a heating block. The enzyme is inactivated at 80°C for 15 min in a heating block. Verify that the plasmid has been completely digested by running an aliquot of the digestion (1 μL or 0.2 μg) along with a nondigested plasmid on a 1% agarose gel.
3. To minimize self-ligated vector in the transformation, linearized vector (9.8 μg) is dephosphorylated in a total volume of

100 μL with 10 U of calf intestinal alkaline phosphatase (CIAP) at 37°C for 15 min in a heating block.

4. To inactivate CIAP add 2 μL of EDTA 0.5 M (pH 8), 5 μL of 10% SDS, and 0.5 μL of proteinase K (10 mg/mL) and incubate at 56°C for 30 min in a heating block.
5. Add an equal volume of phenol, vortex thoroughly, and centrifuge at 16,000 × *g* at room temperature for 15 min. Carefully remove the upper (aqueous) layer and transfer it to a fresh microfuge tube. Add an equal volume of phenol:chloroform: isoamyl alcohol (25:24:1), vortex thoroughly, and centrifuge at 16,000 × *g* at room temperature for 15 min. Carefully remove the upper layer and transfer it to a new microfuge tube.
6. Add 5 M sodium chloride (0.2 M final concentration) and 2 volumes of ice-cold 100% ethanol and store 15 min on ice. Vortex thoroughly and centrifuge at 16,000 × *g* at 4°C for 5 min. Carefully decant off the supernatant.
7. Add 1 mL of 70% ethanol to wash the pellet and recentrifuge at 16,000 × *g* at 4°C for 5 min. Decant off the supernatant and take care not to remove the pellet of DNA. Let the open tube dry on the bench at room temperature for about 20 min.
8. Dissolve the pellet in 100 μL of distilled water.

3.4. Ligation

1. Mix the digested metagenomic DNA (100 ng) and the vector (100 ng) and add 1 μL of T4 DNA ligase in a total volume of 20 μL. For each library, set up independent ligation reactions (5–10) and incubate them at 16°C overnight in a low-temperature water bath.
2. Inactivate the ligase at 65°C for 15 min in a heating block. Transfer all the ligation reactions by pipetting them to a fresh microcentrifuge tube and add water up to a total volume of 200 μL.
3. To precipitate the DNA, add 0.1 volumes of 3 M sodium acetate pH 5.2 and 2.5 volumes of 100% ethanol. Vortex thoroughly and store at –80°C for 1 h. Centrifuge at 16,000 × *g* at room temperature for 15 min. Carefully decant off the supernatant.
4. Add 1 mL of 70% ethanol to wash the pellet and centrifuge at 16,000 × *g* at room temperature for 5 min. Decant off the supernatant without disturbing the pellet of DNA. Let the open tube dry on the bench at room temperature for about 10 min.
5. Dissolve the pellet in 20 μL of distilled water.

3.5. Electroporation

1. Place the DNA (ligation) and the electroporation cuvettes on ice. Store the pipette tips in the freezer until they are used.

2. Thaw an aliquot of Electromax DH10B cells (100 μL) on wet ice.
3. Set the electroporator with the program Ec1 for *E. coli* (0.1cm cuvette) and voltage of 1.8 kV.
4. To electroporate, 2 μL of the ligation are added to the microcentrifuge tube containing the thawed competent cells and mixed gently with pipette. Place the tube again on ice.
5. Pipette the mixture containing cells and DNA into a chilled cuvette. Gently tap cells down to the bottom of the cuvette to ensure that the mixture forms an even layer visible on both sides of the cuvette. Avoid formation of bubbles.
6. Carefully wipe moisture from sides of cuvette and place the cuvette in the electroporator chamber slide. Push the slide until the metal sides of the cuvette are placed firmly between the electrodes. Pulse once and a tone sounds indicates that a pulse has been given. Time constant should be approximately 5.6–6.0 ms.
7. Remove the cuvette from the chamber and immediately add 1 mL of SOC medium to the cuvette. The effect of the time between applying the pulse and transferring the cells to the SOC medium is crucial for cell viability, and a delay in transfer to the outgrowth medium can decrease significantly the transformation efficiency (9). Transfer the cell suspension from the cuvette to a 15-mL culture tube (e.g., Falcon tube). Incubate at 37°C for 1 h, shaking at 225 rpm.
8. Check and record the pulse parameters. Time constant should be 5.6–5.8 ms.
9. Dilute 100 μL of transformed cells 1:20 with SOC medium. The rest (approximately 1 mL) is mixed up to a total volume of 1.5 mL with LB plus 10% (wt/vol) glycerol, vortexed thoroughly, and stored at –80°C.
10. Transfer an adequate volume of the dilution of transformed cells (e.g., 1, 10, and 100 μL) onto LB containing 50 μg/mL ampicillin plates (see Note 3). After transferring cells, plates are left at room temperature for 20–25 min, to allow the absorption of the liquid and subsequently inverted and incubated overnight at 37°C.
11. The next day, count the number of cells grown on plates.

3.6. Plasmid Library Titering

1. An aliquot of the stored library is thawed and placed on ice.
2. Prepare dilutions-3 (1:10^3) and -6 (1:10^6) into a 1.5-mL microcentrifuge tube and mix by gentle vortexing.
3. 1 μL from dilution-3 is transferred to a 1.5-mL microcentrifuge tube containing 50 μL of LB broth and mixed by gentle vortexing. The total volume is spread directly onto a prewarmed LB-Ap plate.

4. 50 and 100 μL aliquots are spread from dilution-6 onto independent LB-Ap plates.
5. The next day, count the number of cells grown on plates. For dilution-3, the cfu/mL is the number of colonies $\times 10^3 \times 10^3$, and for dilution-6, the cfu/mL is (Number of colonies/volume plated) $\times 10^3 \times 10^3 \times 10^3$.
6. Pick 40 single colonies to determine the frequency of transformed cells and to calculate the average size of the insert and its range.
7. Inoculate each colony in a 1-mL culture of LB-Ap in separate 1.5-mL microcentrifuge tubes or in a 96-well flat-bottom block filled in each well with 1.5 mL of LB-Ap.
8. Incubate the cultures overnight (or for 24 h if a 96-well flat-bottom block is used) at 37°C with vigorous shaking (230 rpm).
9. Purify the plasmid with a Qiaprep miniprep kit according to the manual instructions.
10. Digest the plasmids with two restriction enzymes flanking the insertion such as *Eco*RI and *Xba*I. Make up the reaction mixture by adding the reagents into a 1.5-mL sterile microcentrifuge tube: water (up to a total volume of 15 μL), 1 μg of plasmid, 10× buffer, and 0.5 μL of each restriction enzyme. Incubate at 37°C for 1 h in a heat block.
11. Analyze the length polymorphism of the fragments by electrophoresis in a 1% agarose gel run in 1× TBE and reveal by staining with ethidium bromide.

3.7. Library Amplification

1. The library is plated directly on the selective medium. Plate as many cells as necessary (approximately 1.3×10^4 cells per 90-mm plate). Calculate the number of cells from the library titering and use 2–3× the total number of clones in the library. Determine how many plates to use, i.e., if the library has 130,000 independent clones the total number of plates to use will be $130{,}000 \times 2/13{,}000 = 20$ plates. Calculate the volume of medium needed to spread 100 μL on each plate, i.e., 20 plates × 100 μL = 2 mL. As the library aliquots are 1.5 mL, add the necessary volume of LB-Ap up to a total volume of 2 mL in a 15-mL tube and spread 100 μL onto each of the 20 plates of LB-Ap.
2. Grown cells from each plate are scraped up and mixed with 3.5 mL of LB plus 10% (wt/vol) glycerol with a sterile inoculation spreader, pooled in a flask with cells from the same library, and mixed again. Divide the culture in 1 mL aliquots and store them at −80°C.
3. Titre the amplified library as indicated above in Plasmid Library Titering.

3.8. Determination of the Minimal Inhibitory Concentration

1. Prepare a stock solution of the metal at 1 M in distilled water and sterilize by filtration. From this solution, prepare either 100 or 200 mM dilutions to prepare the plates. These solutions can be stored at room temperature away from light.
2. To prepare the plates at different concentrations of the metal mix the LB-Ap and the adequate volume of metal in a flask, mix thoroughly, and dispense approximately 25 mL onto the plates. Let the plates dry for 15–20 min at room temperature (see Notes 4 and 5).
3. To determine the MIC perform a drop assay by inoculating a single colony of DH10B in a 5-mL culture of LB-Ap.
4. Incubate the cultures overnight at 37°C with shaking (230 rpm) in a water bath.
5. Carry out serial dilutions ($0–10^{-4}$) of the culture grown overnight and adjust to 2 the absorbance at 600 nm (Abs 600 nm).
6. Pipette 5 μL of each dilution ($0–10^{-4}$) directly onto the plates with LB-Ap and containing the metal at different concentrations. Incubate the cultures for 16 h at 37°C.
7. The minimum inhibitory concentration (MIC) is the lowest concentration of metal that inhibits the visible growth of the cells. Photograph the plates.

3.9. Screening of the Metagenomic Libraries

1. Thaw an aliquot of an amplified metagenomic library containing approximately 1×10^8 cells. Calculate the volume of medium (LB-Ap) needed to spread 100 μL on each plate.
2. Spread the cells onto plates with LB-Ap containing a metal concentration slightly higher than the MIC. Incubate the cells at 37°C for 24 h or until visible cells are growing on the plates.
3. The next day, count all the visible cells growing on the plates. To exclude chromosomal mutations, the plasmids from all the colonies will be transformed again in new DH10B cells without metal selection. Thus, the colonies are pooled by adding 4 mL of LB-Ap on the plate, and homogenized with a glass sterile inoculation spreader.
4. Purify the plasmid with a Qiaprep miniprep kit according to the manual instructions.
5. Transform chemically competent DH10B *E. coli* cells with 50 ng of the purified plasmid. Mix the contents gently. Place the tubes on ice for 30 min.
6. Place the tubes with cells and DNA in a preheated 42°C water bath for exactly 90 s without shaking.
7. Let the tubes chill for 1–2 min on ice.
8. Add 1 mL of SOC medium to the tube and incubate the cells for 1 h at 37°C with shaking (230 rpm).

9. Transfer an adequate volume (up to 200 μL per plate) of transformed cells onto LB-Ap plates.
10. The next day patch approximately 200 transformed cells onto LB-Ap plates. Number each patch at the bottom of the plate so that the cells can be identified later.
11. Transfer the grown cells to LB-Ap plates containing the metal by patching the cells. Identify each patch with the same number used in the plates without metal. Store the patches from the plates without metal at 4°C so that they can be used later.
12. The next day, count the number of metal-resistant clones and use their corresponding patches from the plates without metal to purify their plasmid and store as a glycerol stock from a single colony.
13. To determine the number of unique metal-resistant clones, digest the plasmid of the selected clones with *Eco*RI and *Xba*I as described above. Analyze the length polymorphism of the fragments by electrophoresis in a 1% agarose gel run in 1× TBE and reveal by staining with ethidium bromide. Those clones whose plasmids display different restriction patterns are selected to accomplish a drop assay as described above in the determination of the MIC along with a DH10B (pSKII$^+$) strain used as a negative control (see Note 6). An example of a drop assay is shown in Fig. 1a.

3.10. Analysis of Metal Resistance Clones

1. The inserts of cloned metagenomic DNA are sequenced on both strands by using the ABI PRISM dye terminator cycle-sequencing ready-reaction kit and an ABI PRISM 377 sequencer, according to the manufacturers' instructions.
2. The sequences of the inserts from metal-resistant colonies are analyzed with the EditSeq, Megalign, and Seqman programs from the DNAStar package.
3. Putative open reading frames (ORFs) are identified using two programs: ORF Finder available at the NCBI web page (http://www.ncbi.nlm.nih.gov/gorf/gorf.html), and Artemis (http://www.sanger.ac.uk/Software/Artemis/). For translation to protein sequences, the bacterial code is selected, allowing ATG, GTG, and TTG as alternative start codons. All the predicted ORFs longer than 90 bp are translated and used as queries in BlastP. The sequences with significant matches are further analyzed with rpsBlast, and their putative function is annotated based on their similarities to COG (Clusters of Orthologous Groups), and Pfam (Protein families). Those sequences with an E value more than 0.001 in the BlastP searches and longer than 300 bp are considered as hypothetical.
4. Transmembrane helices are predicted with TMpred (http://www.ch.embnet.org/software/TMPRED_form.html).

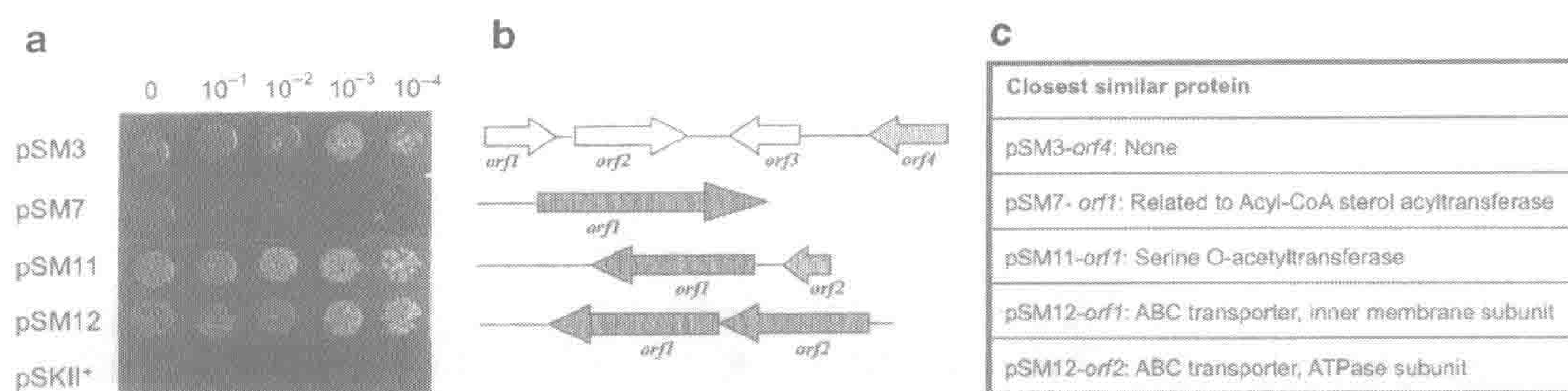


Fig. 1. (**a**) Drop assay of the representative clones pSM3, pSM7, pSM11, and pSM12 obtained from the screening of the metagenomic libraries on LB plates containing 2 mM Ni using serial dilutions ($0–10^{-4}$). *Escherichia coli* strain DH5a (pSKII+) was used as a negative control. (**b**) Schematic organization of the ORFs identified in these plasmids. The *arrows* indicate the locations and the directions of transcription of the ORFs in the different plasmids. Those ORFs involved in nickel resistance are indicated by *gray arrows*. *Arrows* shaded with *vertical bars* indicate the presence of predicted transmembrane helices. (**c**) Closest similar proteins of the amino-acid sequences encoded by the ORFs involved in the resistance phenotype.

5. To determine the ORF(s) involved in metal resistance, two approaches were followed: subcloning and/or in vitro transposon mutagenesis with the Genome Priming System GPS-LS according to manufacturer's instructions (Fig. 1b).
6. Subcloning is performed by PCR amplification using the following reaction mixture: 25 ng of plasmid DNA, 500 μM of each of the four dNTPs, 2.5 U of *Pfu* Turbo DNA polymerase and 100 nM of each forward and reverse primers up to a total volume of 50 μL. The PCR amplification program used is as follows: 1 cycle of 5 min at 94°C, 30 cycles of 30 s at 94°C, 30 s at 52°C, 5 min at 72°C, and finally 1 cycle of 10 min at 72°C. Primer sequences include restriction sites to allow the digestion of the amplification product and subsequent ligation into $pSKII^+$ vector. To include their native expression sequences (promoters and ribosome binding sites), a 200 bp region upstream of the start codon is also amplified. Some of the ORFs are truncated or the 5′ region is close to the polylinker sequence of the $pSKII^+$ vector, and they are subcloned in the same orientation as of the original clone
7. PCR amplification products are gel purified with the Qiaquick Extraction Gel kit and digested overnight with the appropriate restriction enzymes.
8. Digest the $pSKII^+$ vector with the same restriction enzymes for 2–3 h at 37°C.
9. Precipitate both the digested vector and the PCR amplification product with sodium acetate as described for ligation.
10. Ligate both the digested vector and the PCR amplification product overnight at 16°C.
11. Transform the ligation into chemically competent *E. coli* cells as described above.

12. The resulting strains are screened again on LB-Ap containing the metal at the appropriate concentration along with a DH10B ($pSKII^+$) strain used as a negative control.
13. In vitro transposon mutants are transformed in DH10B strain by electroporation and selected on LB-Ap plus kanamycin (20 μg/mL).
14. From each transposition, 200 transformants are patched on LB-agar containing ampicillin plus kanamycin, with and without the metal and grown overnight at 37°C.
15. All mutagenized metal-sensitive and at least ten resistant colonies are rescued from the LB-agar without metal. Metal-sensitive colonies are considered as mutagenized in the genes required for metal resistance, and their plasmids are isolated and sequenced with specific primers for the transposon ends to determine the precise insertion. Plasmids from resistant colonies are also sequenced to localize insertions out of the metal resistance genes.

3.11. Determination of Cellular Metal Concentration

1. *E. coli* DH10B carrying the empty vector and resistant clones are grown in LB liquid medium containing ampicillin at 37°C in a shaking incubator, and growth is measured by absorbance at 600 nm (Abs 600).
2. The metal is added in early stationary phase to the cultures and grown for one additional hour.
3. Cultures are washed four times extensively with ultrapure MilliQ H_2O and centrifugation.
4. Washed pellets are lyophilized, pulverized, and dissolved in H_2O:HCl:HNO_3:H_2O_2 (3:1:4:0.5) by a microwave digestion closed system for subsequent inductively coupled plasma spectroscopy–mass spectrometry (ICP–MS) analysis. An example result is shown in Fig. 2.

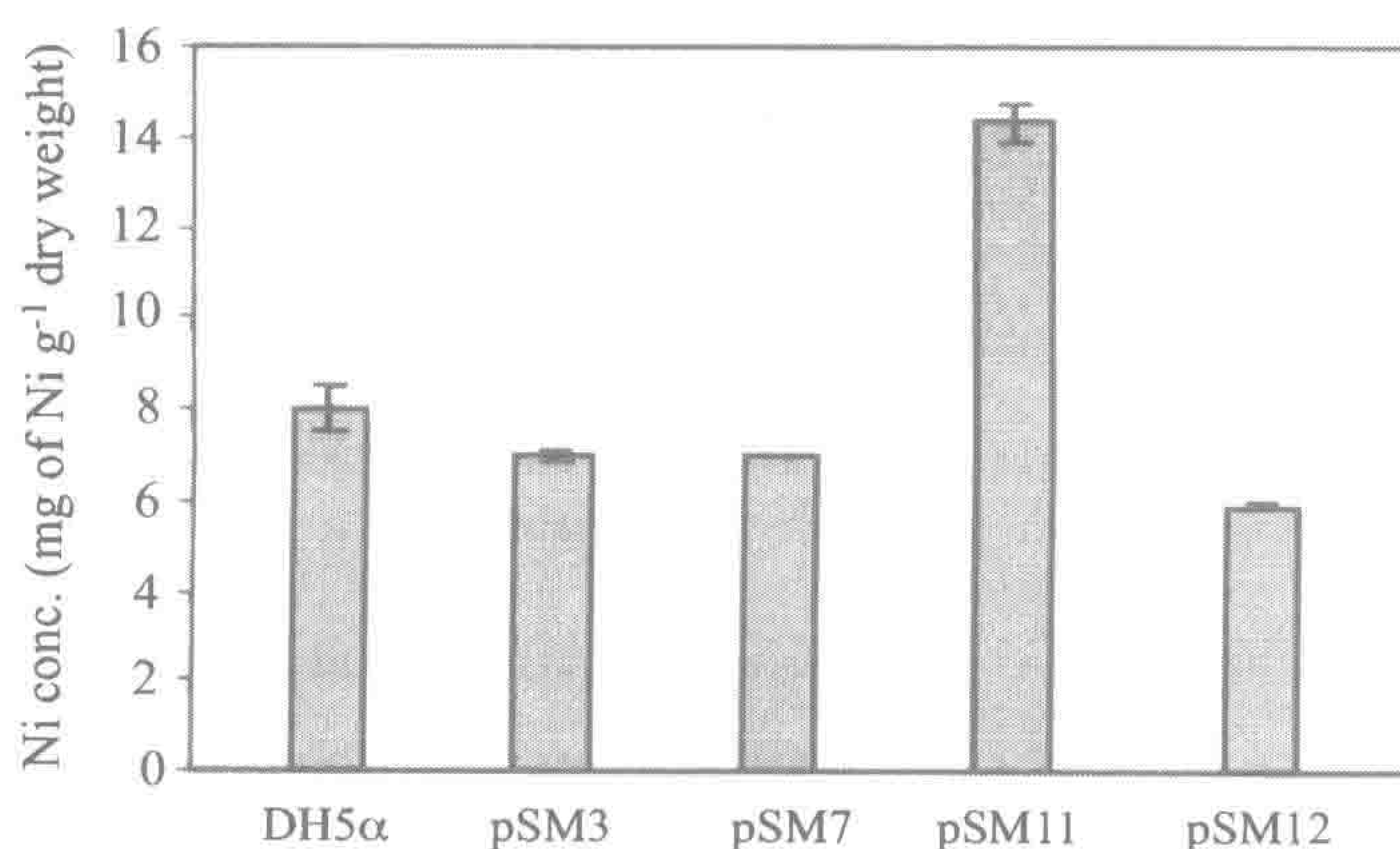


Fig. 2. Test for intracellular content of nickel in the *E. coli* clones pSM3, pSM7, pSM11 and pSM12, and DH5α ($pSKII^+$) with empty vector after 1 h growing with 4 mM Ni. Values are the average of two independent ICP–MS measures. The *error bars* indicate standard deviations.

4. Notes

1. Although the construction of metagenomic libraries can be achieved with the referred kit, other kits and protocols based on direct lysis approaches can be assessed to purify metagenomic DNA from different samples.
2. DNA samples are run in agarose gels prestained with ethidium bromide. *Caution*: Ethidium bromide is a powerful mutagen. Gloves should be worn when handling this product and stained gels. It is advisable to work in a separate place in the lab when handling stained gels. Use appropriate eye and skin protection when observing gels on a UV transilluminator.
3. Unless stated otherwise, cells are plated onto LB containing 50 μg/mL ampicillin (referred to as "LB-Ap" in this text).
4. Some metals are very toxic, and they should be handled with care. Avoid skin contact by wearing gloves and inhalation by using protective masks.
5. For the selection of resistant clones, minimal growth medium such as minimal salts vitamin medium (MSV) (10) can be used instead of LB to avoid complexation and precipitation of some heavy metals.
6. Drop assay can be performed with different metal ions to assess the cross-resistance of the clones under study.

References

1. Bruins, M.R., Magning, J.P., and Oehme, F.W. (2000) Microbial resistance to metals in the environment. *Ecotoxicol Environ Saf* 45, 198–207.
2. Rouch, D.A., Lee, B.T., and Morby, A.P. (1995) Understanding cellular responses to toxic agents: a model for mechanism-choice in bacterial metal resistance. *J Ind Microbiol* 14, 132–141.
3. Silver, S. (1992) Plasmid-determined metal resistance mechanisms: range and overview. *Plasmid* 27, 1–3.
4. Bhadra, B., Nanda, A.K., and Chakraborty, R. (2006) Inducible nickel resistance in a river isolate of India phylogenetically ascertained as a novel strain of *Acinetobacter junii*. *World J Microbiol Biotechnol* 22, 225–232.
5. Rhee, S.K., Liu, X., Wu, L., Chong, S.C., Wan, X., and Zhou, J. (2004) Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Appl Environ Microbiol* 70, 4303–4317.
6. Trajanovska, S., Britz, M.L., and Bhave, M. (1997) Detection of heavy metal ion resistance genes in gram-positive and gram-negative bacteria isolated from a lead-contaminated site. *Biodegradation* 8, 113–124.
7. Mirete, S., de Figueras, C.G., and Gonzalez-Pastor, J.E. (2007) Novel nickel resistance genes from the rhizosphere metagenome of plants adapted to acid mine drainage. *Appl Environ Microbiol* 73, 6001–6011.
8. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, New York.
9. Dower, W.J., Miller, J.F., and Ragsdale, C.W. (1988) High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res* 16, 6127–6145.
10. Teitzel, G.M. and Parsek, M.R. (2003) Heavy metal resistance of biofilm and planktonic *Pseudomonas aeruginosa*. *Appl Environ Microbiol* 69, 2313–2320.

Chapter 20

Retrieval of Full-Length Functional Genes Using Subtractive Hybridization Magnetic Bead Capture

Tracy Meiring, Inonge Mulako, Marla I. Tuffin, Quinton Meyer, and Donald A. Cowan

Abstract

Numerous gene-specific PCR methods have been developed for the cultivation-independent discovery of novel genes from complex environmental DNA samples. The recovery of full-length genes is, however, technically challenging. Here, we present an efficient and relatively simple approach that combines magnetic bead capture with subtractive hybridization for the rapid and direct recovery of full-length target ORFs. When compared with other PCR-based techniques, a higher degree of specificity is achieved through the use of larger gene fragments during hybridization followed by several high-stringency washes. Together with the recent advances in environmental nucleic acid extraction techniques, this approach should allow for the further exploration of the metagenomic sequence space.

Key words: Metagenomics, Gene discovery, Subtractive hybridization, Magnetic bead capture

1. Introduction

Enzymes such as proteases, lipases, and oxidoreductases are of great importance in the pharmaceutical, detergent, food production, and fermentation industries (1). For this reason, there is a constant demand for enzymes that have novel or superior properties, especially new substrate specificities and high levels of enantioselectivity. Environmental samples containing high microbial diversity are an important source of these novel enzymes (2).

Traditionally, the discovery of novel enzymes was based on the screening of bacteria cultured from environmental samples (3). However, over the past two decades, numerous studies have shown that only a small minority of microbial species from any environmental sample can be cultured (3), possibly as low as 1%

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668, DOI 10.1007/978-1-60761-823-2_20, © Springer Science+Business Media, LLC 2010

of the total microbial diversity (4–6). The uncultured majority of microbial species then constitute a huge and largely untapped genetic resource (6).

The development of a "tool kit" of metagenomic-DNA-targeted processes has led to the optimization of protocols for the extraction of DNA directly from soil samples. These methods involve the disruption of bacterial cells by mechanical methods (7) or by chemical and/or enzymatic methods (8), and have been reviewed in detail (5, 9). DNA extracted from soil samples contains humic acids that interfere with restriction enzymes and polymerases (10). The addition of compounds such as polyvinylpolypyrrolidone (PVPP) and hexadecylmethyl ammonium bromide (CTAB) during the DNA extraction procedure reduces the copurification of humic acids (8).

The extraction of good-quality DNA from environmental samples and the development of molecular biology techniques such as gene-specific PCR have made it possible to discover numerous novel genes (1). In gene-specific PCR, a fragment is initially amplified using degenerate primers designed against a conserved nucleotide region (11). The amplified gene fragment is then used to retrieve the full-length open reading frame (ORF). PCR-based methods used to recover flanking regions include seminested PCR (12), rapid genomic walking (13), the vectorette™ system (14), TAIL-PCR (15), inverse PCR (16), and splinkerette (17). Some of these PCR-based techniques are described in detail by Hui et al. (18) and have been used to successfully isolate full-length genes from DNA isolated from pure bacterial isolates. However, due to the heterogeneity of metagenomic DNA samples, these techniques are often difficult to implement on community DNA preparations. Here, we report a subtractive hybridization magnetic bead capture method (Fig. 1), adopted from Jacobsen (10), which paves the way for more efficient recovery of full-length functional genes from metagenomic DNA samples (19). In summary, an internal region of the gene of interest is amplified using degenerate primers labeled with biotin and immobilized on streptavidin-coated magnetic beads. This construct (termed the "driver") is then used as a probe to selectively hybridize the target full-length gene from extracted genomic or metagenomic DNA (termed the "tester"). The subsequent amplification of the full-length gene is facilitated by the addition of T7 primer sites ligated to the tester DNA prior to hybridization.

2. Materials

2.1. Genomic DNA Extraction

These materials are for the extraction of high-molecular weight DNA using the detergent/enzyme-based protocol of Zhou and coworkers (8).

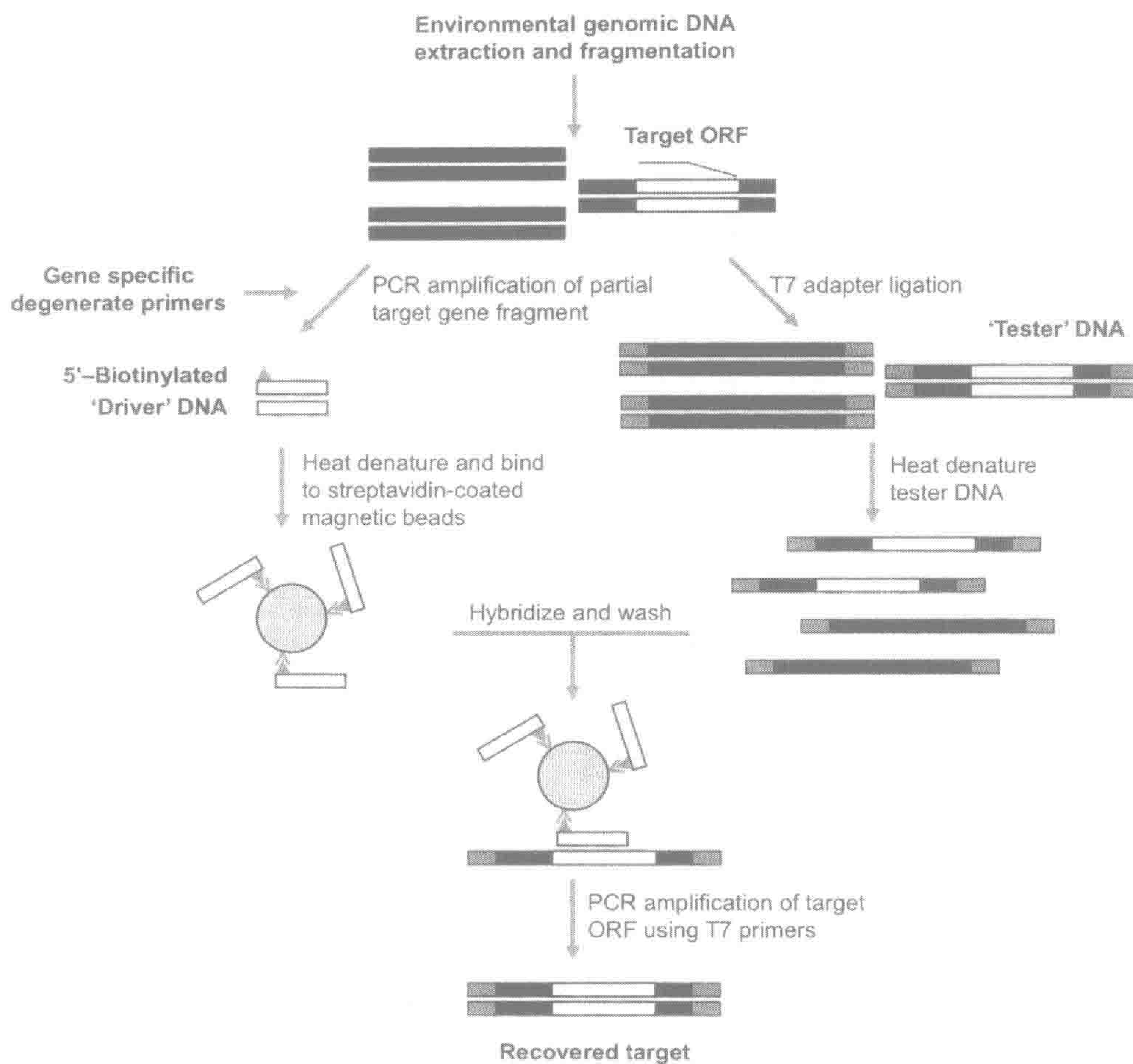


Fig. 1. Schematic outline of the subtractive hybridization magnetic bead capture technique for gene discovery (modified from ref. 19).

1. Extraction buffer: 100 mM sodium phosphate pH 8, 100 mM Tris–HCl, 100 mM EDTA pH 8, 1.5 M NaCl, 1% (w/v) CTAB.
2. 20% SDS.
3. 20 mg/mL Proteinase K.
4. Phenol/chloroform/isoamyl alcohol 25/24/1 (v/v).
5. Chloroform.
6. Isopropanol.
7. 70% (v/v) EtOH (–20°C).
8. TE buffer: 10 mM Tris–HCl, 1 mM EDTA, pH 8.
9. End-It™ DNA End-Repair Kit (Epicentre Biotechnologies).

2.2. Genomic DNA Fragment Preparation

1. Restriction enzyme digestion:
   (a) RsaI.
   (b) Tango buffer: 33 mM Tris–acetate pH 7.9, 10 mM Mg-acetate, 66 mM K-acetate, 0.1 mg/mL BSA (Fermentas).
2. Addition of 3′ A-tail to purified DNA fragments:
   (a) Taq DNA polymerase.
   (b) 10× PCR buffer: 200 mM Tris–HCl pH 8.8, 100 mM KCl, 100 mM (NH4)2SO4, 20 mM MgSO4, 1% (v/v) Triton X-100.
   (c) dATP.
   (d) dNTP mix.

2.3. Agarose Gel Electrophoresis and DNA Purification

1. Agarose (Bioline).
2. 0.5× TBE: 45 mM Tris–HCl, 45 mM boric acid, 1 mM EDTA, pH 8.
3. Ethidium bromide at a final concentration of 0.5 μg/mL.
4. GFX PCR DNA and gel band purification kit (Amersham).

2.4. Tester DNA Preparation

1. T7 adapter Forward and T7 adapter Reverse (see Table 1).
2. 10× annealing buffer: 1 M Tris–HCl pH 7.5, 5 M NaCl, 0.5 M EDTA.
3. Ligation kit (Fermentas).
4. Klenow DNA polymerase, 10× Klenow buffer, and 1 mM dNTP mix.

2.5. Polymerase Chain Reaction and Cloning

1. Taq DNA polymerase.
2. 10× PCR buffer: 200 mM Tris–HCl pH 8.8, 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$, 1% (v/v) Triton X-100.
3. 1 mM dNTP mix.

Table 1
Oligonucleotides used for the production of T7 specific adapters and the T7 primer used for reamplification

Oligonucleotide	Sequence	Tm	Use
T7 Adapter Forward	TTCTAATACGACTCACTATAGGACGC ACCGATACCGATACCGTTACTCGAT	n/a	Priming site introduction
T7 Adapter Reverse	Phos-TCGAGTAACGGTATCGGTGC	n/a	
T7 primer	AATACGACTCACTATAGG	54	Tester reamplification

4. T/A cloning kits: InsT/Aclone™ PCR product cloning kit (Fermentas) or the pGEM®-T easy kit (Promega).

2.6. Biotin-Labeled Gene-Specific Driver Production

1. Streptavidin-coupled Dynabeads® (Dynal, Invitrogen).
2. Magnetic particle separator designed for Eppendorf tubes (Dynal) or an Eppendorf rack with built-in magnet.
3. PBS: 137 mM NaCl, 2.7 mM KCl, 4.3 mM $Na_2HPO_4 \cdot 7H_2O$, 1.4 mM KH_2PO_4, pH 7.3) with 0.1% sodium dodecyl sulphate (SDS).
4. TE buffer: 10 mM Tris-HCl, 1 mM EDTA, pH 8.0 – 1 M NaCl.
5. 0.125 M NaOH – 0.1 M NaCl.

2.7. Subtractive Hybridization of Tester DNA

1. 0.5% SDS in 2× SSC: 0.3 M NaCl, 30 mM sodium citrate, pH 7.
2. 0.1% SDS in 0.1× SSC: 15 mM NaCl, 1.5 mM sodium citrate, pH 7.

2.8. PCR Amplification of Target ORF

1. High-fidelity BIO-X-ACT™ long DNA polymerase system (Bioline).
2. OptiBuffer™ (Bioline).
3. T7 primer (see Table 1).

3. Methods

3.1. Genomic or Metagenomic DNA Fragment Preparation

1. DNA fragments should preferably be in the range of 1–5 kb in size, i.e., large enough to contain full-length target genes but short enough to permit easy reamplification following subtractive hybridization. Good yields of high-molecular-weight DNA are readily obtained using the detergent-based extraction protocol of Zhou et al. (8). DNA (10 μg) is partially digested with 0.5 U *Rsa*I for 15 min at 37°C to yield fragments of a suitable size. The restriction enzyme used, the ratio of enzyme to DNA, and the time period for digestion should, however, be determined in each instance. Alternatively, fragments of appropriate size may be produced with DNA extraction methods that include a mechanical disruption step that randomly shears the DNA. The optimal time and speed at which the mechanical processing occurs can be determined to attain optimal yields of fragments in the suitable size range. For end-repairing of sheared metagenomic DNA, we recommend the End-It™ DNA End-Repair Kit.
2. DNA fragments can be purified from 1% TBE agarose gels (20) using any appropriate purification method (see Note 1).

3. To facilitate the ligation of adapter priming sites, 3′ A overhangs are added to the purified genomic DNA fragments. The terminal transferase activity of Taq DNA polymerase is used for this purpose. The reaction mixture includes approximately 5 μg of DNA, 6 μL 10× PCR buffer, 300 μM dATP, 80 μM dNTP mix, and 1.5 U Taq polymerase in a final volume of 60 μL. Incubate the reaction mixture for 30 min at 72°C. The 3′ A-tailed DNA can then be recovered using the GFX DNA and gel band purification kit, and eluted in a small volume (10–12 μL) of ddH_2O.

3.2. Adapter Oligonucleotide Reconstruction and Ligation

Adapters are ligated to the fragmented metagenomic DNA and include T7 priming sites that allow for the sequence independent reamplification of fragments following subtractive hybridization. Following the addition of the adapters, the adapted metagenomic DNA is referred to as the "tester" DNA.

1. Adapters are produced by annealing two overlapping oligonucleotides, termed T7 adapter Forward and T7 adapter Reverse (see Table 1 and Note 2). Mix 200 nmol of each adapter oligonucleotide and 5 μL 10× annealing buffer in a total volume of 50 μL. Heat the reaction to 94°C for 5 min and then keep at 65°C for 10 min before cooling to 25°C for 1–2 h (see Note 3).
2. Reconstructed adapter DNA is ligated to the fragmented metagenomic DNA (from Subheading 3.1) in a 1:10 molar ratio. Ligation reactions are prepared according to the manufacturer's instructions and incubated overnight at 4°C (see Note 4).
3. Add 2 U Klenow DNA polymerase and 5 μL dNTP mix (1 mM stock) to the ligation mixture and incubate for 30 min at 37°C. Blunt-ending the adapter 5′ overhang is necessary to reconstruct the T7 priming site on the adapter. The tester DNA is purified directly from the reaction using the GFX PCR DNA and Gel Band Purification kit (see Note 5).

3.3. Consensus Gene-Specific Primer Design

As with any screening strategy that is sequence dependent, i.e., PCR-based, the design of consensus gene-specific primers is crucial to success. Primer design is typically a balance between specificity and efficiency, with the aim of targeting and amplifying as many homologous gene fragments as possible with the least amount of concomitant nontarget-specific amplification (reviewed in ref. 21). Primers that are highly specific may severely limit the number of targets or homologous gene fragments recovered, while highly degenerate primers will bind nonspecifically, leading to the amplification of unwanted background or "junk" DNA. The degree of specificity required will depend primarily on the target gene, the identification of appropriate consensus sequences, and the extent of sequence homology across different prokaryotic taxa.

Table 2
Degenerate primers for the recovery of bacterial multicopper oxidases (*mco*)

Primer	Sequence	Tm	Use
Lac1F	GGGCCACGAAACCAGCNTNCAYTGGCA	64	*mco* gene detection
Lac1R	AACCACATGCCGTGCAGRTGNATNGGRTG	64	
Lac1F-5′-Bio	Biotin- GGGCCACGAAACCAGCNTNCAYTGGCA	64	Driver production

In our experiments, we designed consensus sequence-based primers (Table 2) to target multicopper oxidases that successfully amplified homologous genes in a wide array of microbial taxa (Meyer, Burton, and Cowan, unpublished results). This is achieved by aligning the available gene sequences using ClustalW (22) and identifying suitable regions or target sites for degenerate primer design. Forward and reverse degenerate primers are designed to these target sites using the online COnsensus DEgenerate Hybrid Oligonucleotide Primers (CODEHOP) program (23). The parameters used in CODEHOP included a maximum degeneracy of 128-fold and an optimum annealing temperature of 64°C. All other program parameters are set at the default settings.

3.4. Biotin-Labeled Gene-Specific Driver Production

1. An internal region of the gene of interest is amplified using the degenerate primers. The polymerase chain reaction includes 20 ng genomic DNA template in an appropriate volume of 10× PCR buffer with 2 μM degenerate primers, 200 μM dNTPs, and 1 U Taq polymerase.
2. A touchdown PCR is performed with cycling conditions of 94°C for 3 min, 5 cycles of 94°C for 30 s, 64°C for 1 min and 72°C for 1 min, followed by 16 cycles of 94°C for 30 s, 64°C (with a 1°C decrease/cycle) for 1 min, and 72°C for 1 min, then 10 cycles of 94°C for 30 s, 50°C for 1 min, and 72°C for 1 min, and a final extension cycle of 72°C for 5 min. The annealing temperature used (64°C) during PCR was the optimum annealing temperature for the degenerate primers used and would be adjusted for a particular primer set.
3. Amplicons may now be cloned and sequenced to verify primer specificity. Cloning of PCR products can be achieved using any commercially available cloning kit. We used T/A cloning kits; InsT/Aclone™ PCR product cloning kit, or the pGEM®-T easy kit.
4. Biotinylated gene-specific DNA is generated by amplifying the cloned target gene fragments using a 5′-biotinylated gene-specific forward degenerate primer and an unlabeled

reverse degenerate primer in a standard PCR amplification. Biotinylated driver DNA can be produced from a previously cloned gene fragment (for the isolation of a specific ORF from the metagenome) or by direct PCR amplification from metagenomic DNA for the shotgun isolation of multiple ORFs. The amplified biotinylated PCR product is termed the driver DNA.

5. Biotinylated driver DNA is purified from agarose gels using the GFX PCR DNA and Gel Band Purification kit (see Note 6) and denatured by incubating for 3 min at 94°C.
6. Streptavidin-coupled Dynabeads® are washed three times with 500 μL of 1× PBS with 0.1% SDS (pH 7.3) (see Note 7), followed by a single wash in 500 μL of Tris–EDTA (TE) – 1 M NaCl (pH 8.0) and resuspend in 5 μL TE – 1 M NaCl. Following each wash step, the beads are collected using a magnetic particle separator designed for Eppendorf tubes (see Note 8).
7. Add 10 μg biotinylated denatured driver DNA (from step 5) to the washed streptavidin-coated magnetic beads to a final volume of 50 μL and incubate for 60 min at room temperature with agitation.
8. Unbound driver DNA is then removed through successive three wash steps with 500 μL of TE-1 M NaCl (see Notes 9 and 10).
9. The beads are resuspended in 250 μL 0.125 M NaOH – 0.1 M NaCl and incubated at room temperature for 15 min on a mixing platform to denature the bound driver DNA.
10. Wash the beads three times as above to remove the NaOH, resuspend in 10 μL ddH_2O, and use immediately for the subtractive hybridization of the tester DNA.

3.5. Subtractive Hybridization of Tester DNA

The subtractive hybridization protocol described here was adopted from Jacobsen and coworkers (10).

1. Tester DNA (5 μg, from Subheading 3.2) is added to 10 μL of the beads prepared as in Subheading 3.4 above to a final volume of 100 μL. Hybridization is performed at 68°C to ensure maximum specificity and incubated overnight with continuous shaking.
2. All unbound tester DNA is then removed by washing for 5 min at room temperature in 0.5% SDS, 2× SSC, followed by a further 5 min wash at 68°C in 0.1% SDS, 0.1× SSC (see Note 11).
3. The washed magnetic beads are recovered, resuspended in 10 μL ddH_2O, and stored at 4°C until further use.

3.6. PCR Amplification of Target ORF

To avoid the introduction of random mutations during reamplification, a high-fidelity DNA polymerase capable of amplifying long DNA fragments, such as the BIO-X-ACT™ long DNA polymerase, is recommended.

1. The magnetic beads with the captured or immobilized tester DNA (from Subheading 3.5) can be used directly in PCR for reamplification of full-length target genes. No dissociation step is required; typically, 2 μL beads were added per 50 μL PCR reaction mix.
2. For the PCR, add a final concentration of 1× OptiBuffer™, 2.5 mM $MgCl_2$, 200 μM dNTP mix, 0.4 U BIO-X-ACT™ – long polymerase, and 0.5 μM T7 primer (see Table 1) to the magnetic beads.
3. Cycling conditions are as follows: 94°C for 3 min followed by 30 cycles of 94°C for 30 s, 54°C for 1 min, and 72°C for 3 min with a final extension cycle of 72°C for 15 min.

4. Notes

1. We have found the GFX PCR DNA and Gel Band purification kit (Amersham) to be simple with minimal loss of DNA during extraction.
2. Note that the T7 Adapter Reverse primer is the partial complement of the T7 Adapter Forward and must be phosphorylated at the 5′ end.
3. Reconstructed adapters can be stored indefinitely at –20°C.
4. With complex metagenomic DNA samples, the use of a high-ligase concentration (400 U per reaction), as recommended for adapter ligation (20), gave the best results in our hands.
5. The effective ligation of adapter priming sites to the genomic DNA can be verified by PCR amplification of 20 ng of the tester DNA using the T7 primer (Table 1, see Subheading 3.6).
6. The purification of biotinylated PCR products is not strictly necessary, but if any excess biotinylated primer is present, it will reduce the binding ability of the beads.
7. The beads should be washed thoroughly to remove the NaN_3 that was used as a preservative.
8. A small amount of the streptavidin-coupled beads will be lost during the washing processes as some nonspecific attachment to pipette tips and mirocentrifuge tubes occurs; we, therefore, recommend that the initial amount of beads used should be in excess. Alternatively, siliconized pipette tips and microcentrifuge tubes can be used to reduce the loss of beads.

9. As a measure of the effectiveness of the binding of the driver DNA to the beads, a sample of the supernatant should be kept and run on an agarose gel or the OD-reading checked to determine the concentration of unbound driver.
10. Removal of unbound driver DNA is critical as the tester DNA will bind to any excess free driver molecules present which will decrease the amount of tester hybridized to the beads and subsequently recovered.
11. This is a critical step to remove nonspecific hybridization and prevent background amplification.

References

1. Ogawa, J. and Shimizu, S. (1999) Microbial enzymes: new industrial applications from traditional screening methods. *Trends Biotechnol* 17, 13–21.
2. Lorenz, P. and Eck, J. (2005) Metagenomics and industrial applications. *Nat Rev Microbiol* 3, 510–516.
3. Amann, R.I., Ludwig, W., and Schleifer, K.-H. (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* 59, 143–169.
4. Schloss, P.D. and Handelsman, J. (2003) Biotechnological prospects from metagenomics. *Curr Opin Biotechnol* 14, 303–310.
5. Handelsman, J., Liles, M., Mann, D., Riesenfeld, C.S., and Goodman, R.M. (2002) Cloning the metagenome: culture independent access to the diversity and functions of the uncultured microbial world. In: Wren, B. and Dorrell, N. Eds. Functional Microbial Genomics. Academic, New York, pp. 241–255.
6. Cowan, D., Meyer, Q., Stafford, W., Muyanga, S., Cameron, R., and Wittwer, P. (2005) Metagenomic gene discovery: past, present and future. *Trends Biotechnol* 23, 321–329.
7. Bürgmann, H., Pesaro, M., Widmer, F., and Zeyer, J. (2001) A strategy for optimizing quality and quantity of DNA extracted from soil. *J Microbiol Methods* 45, 7–20.
8. Zhou, J., Bruns, M.A., and Tiedje, J.M. (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbiol* 62, 316–322.
9. Pettit, R.K. (2004) Soil DNA libraries for anticancer drug discovery. *Cancer Chemother Pharmacol* 54, 1–6.
10. Jacobsen, S.C. (1995) Microscale detection of specific bacterial DNA in soil with a magnetic capture-hybridization and PCR amplification assay. *Appl Environ Microbiol* 61, 3347–3352.
11. Eschenfeldt, W.H., Stols, L., Rosenbaum, H., Khambatta, Z.S., Quaite-Randall, E., Wu, S., et al. (2001) DNA from uncultured organisms as a source of 2,5-diketo-D-gluconic acid reductases. *Appl Environ Microbiol* 67, 4206–4214.
12. Sheu, D.-S., Wang, Y.-T., and Lee, C.-Y. (2000) Rapid detection of polyhydroxyalkanoate-accumulating bacteria isolated from the environment by colony PCR. *Microbiology* 146, 2019–2025.
13. Kilstrup, M. and Kristiansen, K.N. (2000) Rapid genome walking: a simplified oligo-cassette mediated polymerase chain reaction using a single genome-specific primer. *Nucleic Acids Res* 28, E55.
14. Aarnikunnas, J., Rönnholm, K., and Palva, A. (2002) The mannitol dehydrogenase gene (*mdh*) from *Leuconostoc mesenteroides* is distinct from other known bacterial *mdh* genes. *Appl Microbiol Biotechnol* 59, 665–671.
15. Huang, H., Luo, H., Yang, P., Meng, K., Wang, Y., Yuan, T., et al. (2006) A novel phytase with preferable characteristics from *Yersinia intermedia*. *Biochem Biophys Res Commun* 350, 884–889.
16. Uchiyama, T. and Watanabe, K. (2006) Improved inverse PCR scheme for metagenome walking. *Biotechniques* 41, 183–188.
17. Too, W.C., Liew, Y.C., and Few, L.L. (2008) Cloning of glyceraldehyde-3-phosphate dehydrogenase from an Antarctic psychrophilic bacterium by inverse and splinkerette PCR. *J Basic Microbiol* 48, 430–435.
18. Hui, E.K., Wang, P.C., and Lo, S.J. (1998) Strategies for cloning unknown cellular flank-

ing DNA sequences from foreign integrants. *Cell Mol Life Sci* 54, 1403–1411.

19. Meyer, Q.C., Burton, S.G., and Cowan, D.A. (2007) Subtractive hybridization magnetic bead capture: a new technique for the recovery of full-length ORFs from the metagenome. *Biotechnol J* 2, 36–40.
20. Sambrook, J. and Russell, D.W. (2001) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Press, New York.
21. Linhart, C. and Shamir, R. (2005) The degenerate primer design problem: theory and applications. *J Comput Biol* 12, 431–456.
22. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22, 4673–4680.
23. Rose, T.M., Schultz, E.R., Henikoff, J.G., Pietrokovski, S., McCallum, C.M., and Henikoff, S. (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res* 26, 1628–1635.

Chapter 21

Detection and Isolation of Selected Genes of Interest from Metagenomic Libraries by a DNA Microarray Approach

Gopal P. Pathak and Wolfgang Gärtner

Abstract

A DNA microarray-based approach is described for screening metagenomic libraries for the presence of selected genes. The protocol is exemplified for the identification of flavin-binding, blue-light-sensitive biological photoreceptors (BL), based on a homology search in already sequenced, annotated genomes. The microarray carried 149 different 54-mer oligonucleotides, derived from consensus sequences of BL photoreceptors. The array could readily identify targets carrying 4% sequence mismatch, and allowed unambiguous identification of a positive cosmid clone of as little as 10 ng against a background of 25 μg of cosmid DNA. The protocol allows screening up to 1,200 library clones in concentrations as low as ca. 20 ng, each with a ca. 40 kb insert size readily in a single batch. Calibration and control conditions are outlined. This protocol, when applied to the thermophilic fraction of a soil sample, yielded the identification and functional characterization of a novel, BL-encoding gene that showed a 58% similarity to a known, BL-encoding gene from *Kineococcus radiotolerans* SRS30216 (similarity values refer to the respective LOV domains).

Key words: Blue-light photoreceptor, LOV domain, Metagenome screening, Microarray

1. Introduction

A DNA microarray-based approach allows convenient screening of metagenomic DNA libraries that generally contain thousands of different clones. The heterologous expression of large-insert clones and the need for the correct assay to detect the functional expression are most common challenges that one faces during metagenomic DNA library screening. A most topical application of the microarray technology is its use as a tool in environmental profiling of (potentially) functional genes (1, 2). Clearly, the similarity of novel DNA sequences to already identified genes encoding for functional proteins is the basis for this approach.

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668, DOI 10.1007/978-1-60761-823-2_21, © Springer Science+Business Media, LLC 2010

In most metagenomic screening approaches, however, such proposed protein function has not been demonstrated, be it for the nonavailability of sufficient amounts of DNA or the missing knowledge of environmental or cellular factors that regulate the activity of the identified gene product. It is, however, possible to extend the information derived from a DNA microarray screening in the direction of a protein function, as is outlined here, exemplifying the search and identification of a novel blue-light (BL)-sensitive photoreceptor protein (3).

Light in general and the blue-light fraction of the visible spectrum in particular is a strong environmental signal to which all organisms have to respond, either in a tactic manner or via the regulation of gene expression. The rationale for an instant response is probably the most threatening situation that blue light generates with a high quantum yield the triplet state of porphyrins, ubiquitous in all living organisms, a reaction, which in turn produces singlet oxygen as a most reactive and deleterious molecule. The largest class of photoreceptors that detect the blue-light spectral range contain a LOV (light, oxygen, voltage) domain that functions via an incorporated flavin chromophore (λ_{max}=447 nm) (4, 5). Irradiation with light of a potentially harmful spectral range has to be detected and counteracted by a variety of physiological responses. It is, thus, not surprising that ca. 20% of all sequenced microbial genomes carry one or even several genes coding for BL-sensitive photoreceptors (6). It is, thus, suggestive to investigate metagenomes for the presence of – potentially – novel genes that encode BL-sensitive photoreceptors, in particular, since microarrays, constructed from synthetic oligomers corresponding to the conserved motif of similar functional genes have been shown to detect genes containing similar motifs (7). Extending its scope, the DNA microarray technique has been successfully applied to detect and isolate a functional novel gene encoding a BL photoreceptor from a metagenome (3).

In this approach, we present the employment of a microarray, based on 149 spotted oligonucleotides, for the search of BL-sensitive photoreceptor-encoding genes in a metagenome of a soil sample (in this application, the thermophile fraction was screened that was obtained by incubating soil at 65°C for 6 months). DNA material from this preparation was cloned into a cosmid library, allowing to screen in each batch thousands of clones in short time. After a positive signal was detected, the initial number of clones in the pool was reduced in consecutive screenings. After the final positive cosmid clone had been identified, its large, ca. 40 kb insert was subcloned into plasmids aiming at the identification of the target gene. This approach yielded a BL-coding gene of 3,615 nucleotides, corresponding to a 1,204 amino acid protein with a molecular weight of 135 kDa (3). The predicted multidomain protein (a property quite typical for prokaryotic BL-photoreceptors (6)) contained four PAS domains, out of

which one turned out to be the flavin chromophore-bearing LOV domain (Fig. 2). In addition, two motifs, one for a histidine kinase and another one for a response regulator, as found in many sensory two component signaling systems (8), are present in this putative BL photoreceptor (Fig. 1).

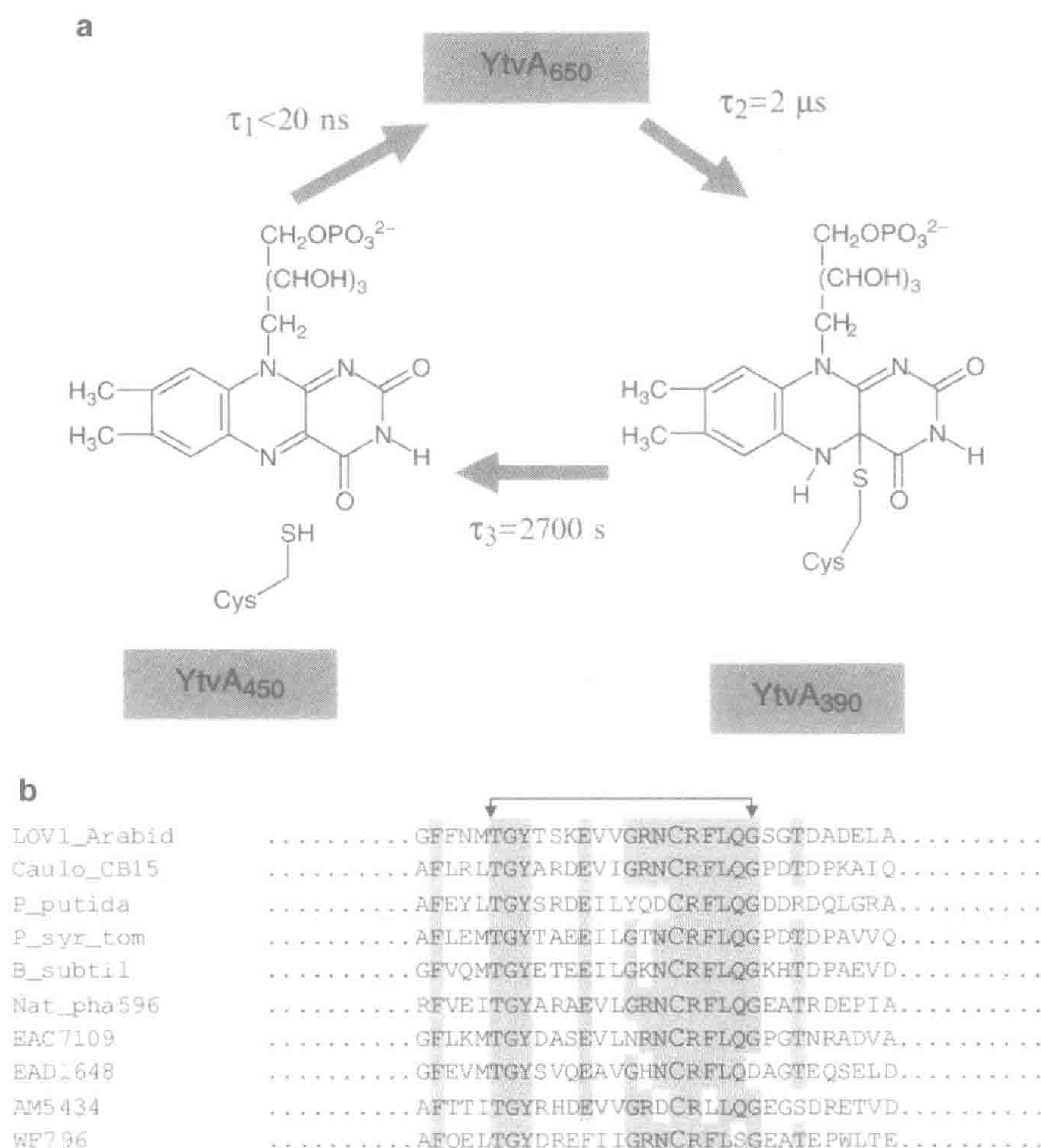


Fig. 1. (**a**) Photocycle of a LOV domain-protein (YtvA from *B. subtilis*). Blue light irradiation forms the triplet state of the horizontal bar limited by arrowheads FMN (flavin mononucleotide) chromophore ($YtvA_{650}$). During decay of the triplet state (τ ca. 2 μs), a covalent bond to a nearby positioned cysteine residue is formed. Upon dark incubation, this covalent bond reopens to form the parent state in minutes or seconds (depending on the origin of the protein). (**b**) Sequence alignment of the core region exemplified for ten LOV domains. Indicated is the range of 18 amino acids that were used to design 54-mer oligonucleotides that were spotted for the microarray approach. The fully conserved, functionally essential cysteine (photoadduct formation) is enlarged; LOV1_Arabid: *Arabidopsis thaliana* Phot1-LOV1; Caulo_CB15: *Caulobacter crescentus* CB15; P_putida: *Pseudomonas putida* KT2440, SB1; P_syr_tom: *Pseudomonas syringae* pv. *tomato*; B_subtil: *Bacillus subtilis*; Nat_pha596: *Natronomonas pharaonis* DSM2160; EAC7109: Sargasso metagenome EAC77109; EAD1648: Sargasso metagenome EAD41648; AM5434: Acid mine drainage metagenome (gi: 41580434); WF796: Whale fall metagenome (gi: 60178796). The horizontal bar limited by arrowhead indicates the region of highest conservation being used for oligonucleotide design for microarray construction. The covalent-forming cysteine residue is shown in *bold*.

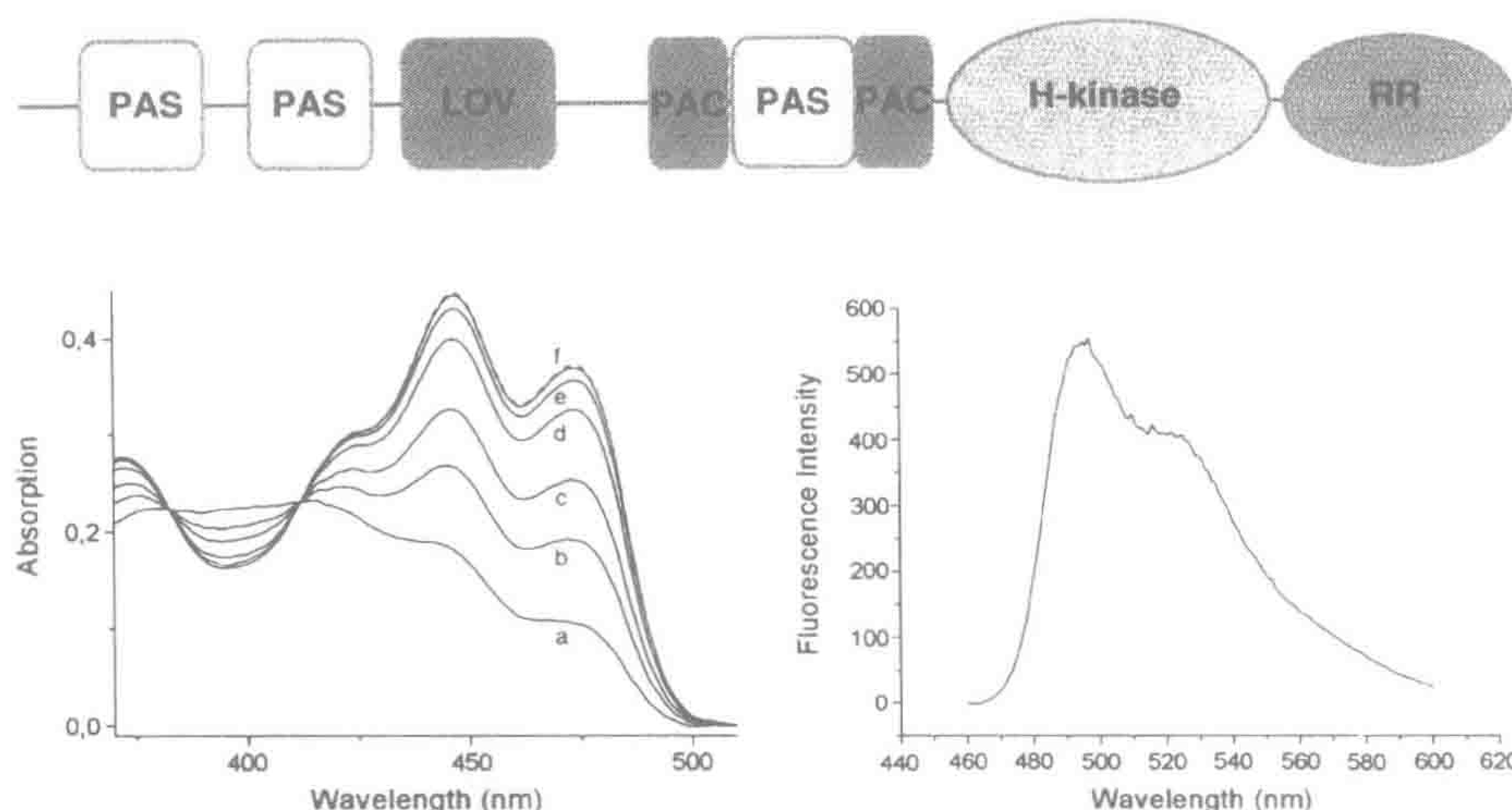


Fig. 2. Properties of metagenome-derived LOV domain protein HT-Met1. Top: Domain structure, PAS, PAC, LOV, protein domains; H-kinase, histidine kinase; RR, response regulator. *Bottom, left*: Photochemistry of recombinant HT-Met1-LOV; the dark state (λ_{max} = 447 nm, dashed curve) shows the characteristic three-peaked absorption band of an oxidized, protein-bound flavin. Curve a: maximal generation of lit state by continuous blue light irradiation; curves b–f (corresponding to selected time points of the dark state recovery process), 80 s, 2 min, 4 min, 8 min, 12 min after irradiation, respectively (T = 20°C). The lifetime of the dark recovery reaction was ca. 120 s. Bottom, right: Fluorescence spectrum of the HT-Met1 LOV domain (λ_{ex} = 450 nm).

Subcloning of the chromophore-binding LOV domain indeed yielded a photoactive protein with all properties significant for a prokaryotic phototropin-type BL photoreceptor (Note 1) (9). The photochemistry is demonstrated in Fig. 2.

2. Materials

2.1. Microarray Hybridization

1. 5′-End amino modified oligonucleotide probes (54-mer) (Operon Biotechnologies) spotted on CodeLink™ activated slides (GE Healthcare); store at room temperature inside a desiccator.
2. Wash solution 1: 1× SSC, 0.2% w/v SDS; store at room temperature.
3. Wash solution 2: 0.1× SSC, 0.2% w/v SDS; store at room temperature.
4. Wash solution 3: 0.1× SSC; store at room temperature.
5. Tom Freeman hybridization buffer: 40% formamide, 5× Denhardt's solution, 5× SSC, 1 mM Na-pyrophosphate, 50 mM Tris–HCl pH 7.4, and 0.1% SDS; store at −20°C.
6. Hybdrization station (GeneMachines® Hyb4 from Genomic Solutions, USA) including editable software (Hyb4 editor).
7. Microarray scanner GenePix 4100A scanner (Axon Instruments Inc, CA, USA) equipped with image analysis software (GenePix Pro 6).

8. Table-top centrifuge (Hettich, Germany).
9. In all preparations, tridistilled water and chemicals of analytical grade are used.
10. Routine cloning procedures and medium/solution composition follow the recipes given in ref. (10).

2.2. DNA Library Replication and Maintenance

1. LB medium with required antibiotic supplement.
2. Multichannel pipettes (1,000 and 100 μl, Eppendorf, Germany).
3. 96-Well microplates (Corning, USA or Greiner, Germany).
4. 96-Pin sterilizable metal block for duplication of library clones.

2.3. DNA Isolation, Manipulation, and Cloning

1. 2× LB medium.
2. 2.2 mL culture plate in 96-deep-well format (Eppendorf, Germany).
3. Cosmid/BAC DNA preparation kit in 96-well format (Qiagen, Germany or Favorgen, Taiwan, etc.).
4. 50 mL polypropylene centrifuge tubes (Corning, USA, used only when the pooling method is applied, see below).
5. Swing out rotor for 96-deep-well plate (Hettich, Germany).
6. Beckman Coulter™ Avanti™ J-20 XP centrifuge with JLA 16250 rotor and appropriate adapter for 50 mL polypropylene tubes (Beckman Coulter, USA).
7. Alkaline lysis solution P1, P2, P3, chloroform, isopropanol, ethanol (absolute and 70%) according to ref. (10).
8. Restriction enzymes and DNA-modifying enzymes (Fermentas, Germany).
9. Plasmid cloning vector (pHSG399 or pUC18).
10. QIAquick PCR purification kit (Qiagen, Germany).

2.4. Fluorescent Labeling of DNA

1. Cy5 and Cy3 dCTP (GE Healthcare, Germany); store at –20°C.
2. dATP, dCTP, dGTP, dTTP (GE Healthcare, Germany); store at –20°C.
3. 10× Random hexamers (Roche, Germany); store at –20°C.
4. Klenow polymerase, exo- (Fermentas, Germany); store at –20°C.
5. *Alu*I restriction enzyme (Fermentas, Germany); store at –20°C.
6. 0.5 M EDTA.
7. Illustra™ CyScribe™ GFX™ purification kit (GE Healthcare, Germany).
8. Vacuum concentrator (Speedvac®, USA).

2.5. Dot Blot Hybridization and Colorimetric Detection

1. 5′-DIG labeled oligonucleotide probes (54 bp, Metabion), corresponding to the positive microarray probe.
2. Positively charged nylon membrane (Roche, Germany).
3. Whatman filter paper.
4. Denaturation solution: 1.5 M NaCl, 0.5 M NaOH); store at room temperature.
5. Neutralization solution: 1.5 M NaCl, 0.5 M Tris–HCl, 1 mM EDTA pH 7.2); store at room temperature.
6. Fixation solution: 0.4 M NaOH; store at room temperature.
7. DIG1 buffer: 0.1 M Tris–HCl, 0.1 M NaCl, pH 7.5; store at room temperature.
8. Blocking reagent (Boehringer, Mannheim, Germany), store at 4°C.
9. Blocking solution: 1% blocking reagent in DIG1 buffer, store at 4°C.
10. Hybridization buffer: 5× SSC, 0.1% w/v N-Laurylsarkosin, 0.02% SDS, and 1% w/v blocking reagent, store at 4°C.
11. Wash buffer 1: 2× SSC, 0.1% w/v SDS; store at room temperature.
12. Wash buffer 2: 0.5× SSC, 0.1% SDS w/v; store at room temperature.
13. Antidigoxigenin-alkaline phosphatase (anti-DIG) (Roche, Germany), store at 4°C.
14. Antibody solution: 150 mU/mL anti-DIG in DIG1 buffer, prepare fresh.
15. Detection buffer: 0.1 M Tris–HCl, 0.1 M NaCl, pH 9.5, store at room temperature.
16. NBT/BCIP stock solution (Roche, Germany), store at 4°C.

3. Methods

3.1. Culture Preparation for DNA Isolation

1. Fill each well of a 96-deep-well culture plate with 1.2 mL of 2× LB medium containing the appropriate antibiotic.
2. Thaw a 96 well microtiter plate containing the library stock on ice. Transfer simultaneously inocculant from each well of the microtiter plate to the corresponding wells of a 96 deep-well plate by the help of a 96-pin sterile metal block.
3. Cover the 96 deep-well culture plate with a sterile, semiporous sealing membrane.
4. Incubate from 18 to 22 h at 37°C, shaking at about 200 rpm.

3.2. DNA Isolation from Cosmid Libraries

1. After the incubation, the cosmid DNA from the library can be extracted using homemade alkaline lysis solutions (10) or 96-well BAC/cosmid DNA extraction kits. In order to apply the alkaline lysis procedure, the clones can be pooled in batches of larger volumes (see Note 2). Here, the extracted DNA was obtained from groups of pooled clones.

3.2.1. Pooling of Clones and Alkaline Lysis

1. Pool the culture in different groups into 50 mL polypropylene tubes (i.e., from 24 or 32 wells).
2. Harvest the pooled culture solution by centrifugation at 5,400 rcf (=×*g*).
3. Apply the alkaline lysis procedure using the homemade lysis solutions (10).
4. Store the extracted DNA at 4°C till further use or at −20°C for longer time.

3.2.2. Kit Based Extraction

1. 96 Well BAC or cosmid DNA isolation kit can be used to extract the cosmid or BAC DNA from the metagenomic library. The liquid culture "from a" 96-deep-well culture plate is harvested by centrifuging at 2,200 rcf (=×*g*) and subjected to the lysis procedure outlined as in the kit suppliers' manual. Elution-based or precipitation-based kits are available. Elution-based kits are preferred over precipitation-based 96-well DNA extraction kits, as the latter ones are more prone to the loss of DNA during precipitation and washing steps. Kit-supported handling is relatively costly and at that step and with amounts of several hundred cosmid clones, it is not required to apply.

3.3. Microarray Probe Preparation

The nucleotide sequences corresponding to the highly conserved and functionally important amino acids are used to prepare the oligonucleotide probes. A protein motif containing 18 highly conserved amino acid residues was selected (Fig. 1), and the corresponding nucleotide sequences were identified from the NCBI database. The corresponding 54 base-5′ end-amino modified oligonucleotide probes were synthesized (Operon Bio-techno-logies) and spotted in quadruplicate on CodeLink™ activated glass slides at the Department of Genomic and Applied Microbiology, University of Göttingen.

3.4. Target DNA Preparation for Microarray Hybridization

1. Aliquots from about 400 to 1,000 clones, each with a concentration of about 20–50 ng, are mixed in an Eppendorf tube.
2. The DNA is concentrated by precipitation with absolute ethanol, pelleted by centrifugation, washed with 70% ethanol, and finally dissolved in H_2O.
3. The concentrated target DNA is digested completely using *Alu*I (4 bp sequence-recognizing restriction enzyme).

4. Digested DNA is purified using the QIAquick PCR purification kit. After elution, the DNA is concentrated by vacuum centrifugation at room temperature. Alternatively, the phenol/chloroform extraction procedure can be applied.
5. In an Eppendorf tube, adjust the volume of digested and purified DNA to 30 μl with H_2O; add 5 μl of 10× random hexamer and 5 μl of Klenow buffer to it. Heat at 96°C for 5 min to denature, snap chill on ice.
6. On ice, add 5 μl of 10× dNTP mix (1.2 mM each dATP, dGTP, and dTTP and 0.6 mM dCTP).
7. Add 2.5 μl of Cy5- or Cy3-dCTP.
8. Add 10–20 units of Klenow fragment (exo-).
9. Adjust the final reaction volume to 50 μl.
10. Incubate at 37°C for 3 h or overnight and then stop the reaction by addition of 5 μl of 0.5 M EDTA.
11. Purify the labeling reaction using the illustra™ CyScribe™ GFX™ purification kit and elute the labeled DNA with H_2O.

3.5. Microarray Hybridization

1. Concentrate the eluted, labeled DNA in the vacuum centrifuge and adjust the volume with H_2O to 30 ml.
2. Denature the labeled DNA at 95°C for 5 min and snap chill on ice.
3. Add Tom Freeman Hybridization buffer to the labeled DNA sample (three times the volume, e.g., 90 μl of hybridization buffer to 30 μl of labeled DNA).
4. Load the microarray slide into the hybridization chamber.
5. Preheat the sample to the initial hybridization temperature, preheat the hybridization station at 72°C for 2 min.
6. Introduce the sample into the hybridization station at the initial hybridization temperature (65°C), hybridize the microarray using step-down temperature programs: 4 h at 65, 60, 55, and 50°C each.
7. Wash slides as follows:

   Two cycles with wash buffer 1 at 45°C, flow for 30 s, and hold for 30 s.

   Two cycles with wash buffer 2 at 42°C, flow for 30 s, and hold for 30 s.

   Two cycles with wash buffer 3 at 42°C, flow for 30 s, and hold for 30 s.

   Drain the slides for 40 s and then dry the slides by spinning at 600 rpm for 10 min.

3.6. Scanning Hybridized Microarray Slides

1. Select the hardware-setting menu from the scanner software and select both lasers to scan at both (635 and 532 nm) wavelengths.
2. To obtain a good signal by avoiding pixel saturation, PMT (photomultiplier tube) setting of the scanner should be adjusted at the beginning; the suggested red and green signal pixel ratio is approximately 1.0, which should be uniformly maintained for all slides (alternatively the automatic PMT balance option can be used).
3. Preview-scan (low resolution scan) the slide to find the area with images, select the area of interest and scan the selection in high-resolution "Data Scan" mode.
4. Save the images as 16-bit TIFF files.
5. Superimpose the grid of individual circles, stored in the program as a (*.gal) file, defining and specifying each probe on the microarray from the GenePix array list to the microarray image to assign each fluorescent spot to the related probe.
6. Analyze the image data using the provided software (GenePix Pro 6).
7. Determine the signal intensity of the probe; the following parameters need to be read: SI=F635 mean – B635 "(F, B, fluorescence, background at 635 nm, respectively) for cy5-labeled DNA, F532 mean – B532 for cy3-labeled DNA (SI, signal intensity; F635/532, B635/532, fluorescence, background at 635/532 nm), and the signal to noise ratio (SNR=signal intensity – background/standard deviation of background).
8. A positive hybridization should produce a signal intensity of at least three times higher than the background. A SNR value of more than 5 can be taken as positive hybridization to avoid false positive ones (see Note 3).

3.7. Identification of Positive Clones

After the detection of a positive signal in the pool of clones, the number of clones in the pool is broken down always by half in subsequent experiments (Fig. 3). Starting from the pool of 800 clones, the clones are divided into two pools of 400 clones each and tested again. After the detection of the same positive signal, these 400 clones are divided again into 200 each and so on. Dual labeling using Cy5 and Cy3 dye can be used to reduce the number of hybridization experiments. After few subsequent rounds of hybridizations, the positive clone is determined.

3.8. Identification of a Positive Gene

3.8.1. Subcloning of a Positively Identified Cosmid Clone and Detection of Positive Plasmid Clones

In order to minimize the size of the insert and to allow more convenient handling, the cosmid clone is fragmented by restriction enzymes and cloned in plasmid vectors (Fig. 4). As a positive probe is already determined from the microarray, the same oligonucleotide sequence can be used to synthesize a DIG-labeled probe. The plasmid clones can be hybridized with the DIG-labeled oligonucleotide primer. In this test, a single probe

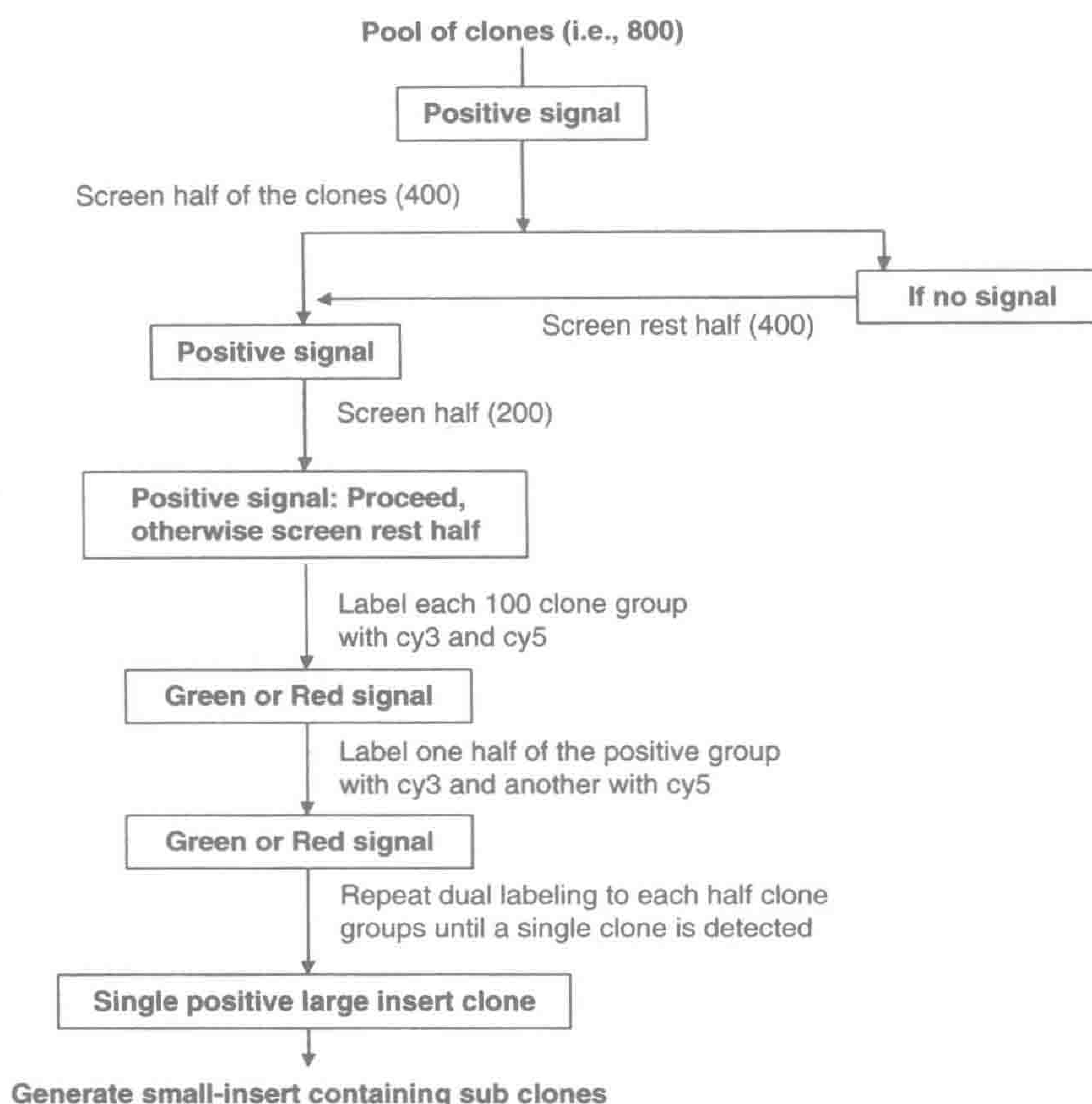


Fig. 3. Outline of metagenomic library screening approach. In the example presented here, the initial batch consisted of 800 clones, which were reduced by half in each scan later. A dual dye approach was also used, which helps to avoid false-positive signals by experimental errors.

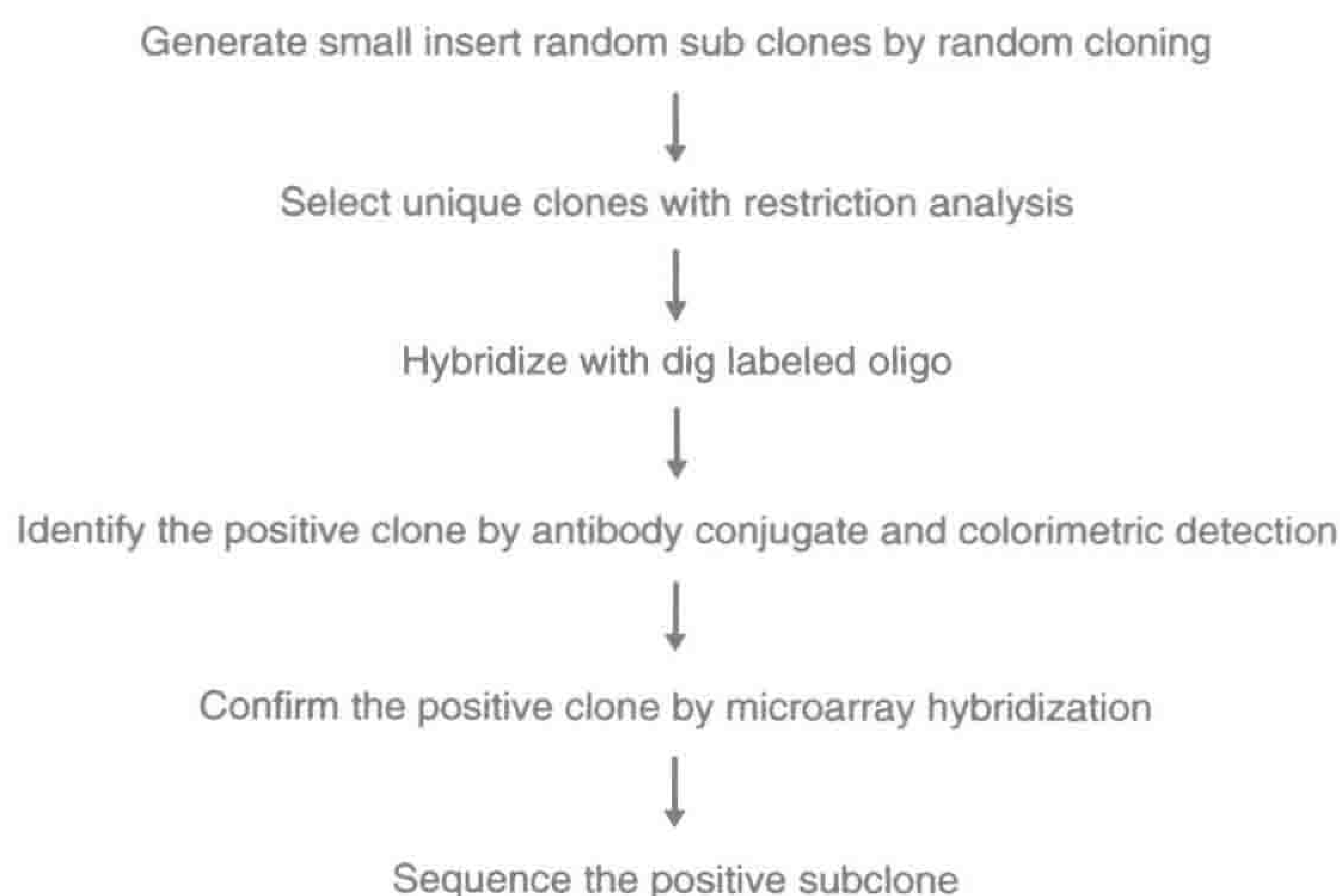


Fig. 4. Brief outline of the steps to detect a subclone for final sequencing.

can be hybridized to a number of individual clones simultaneously, which is more cost-effective than the microarray hybridization and time saving too.

3.8.2. DIG Labeling and DNA Dot Blot Hybridization

3.8.2.1. Dot Blotting of Plasmid DNA to the Membrane

1. Heat-denature plasmid/cosmid DNA at 96°C for 6 min and snap cool in ice for 3–5 min (keep on ice until further use).
2. Place the positively charged nitrocellulose membrane on Whatman filter paper, make grid with pencil to drop the denatured DNA.
3. Apply 5 μl of DNA (about 5–15 ng) on each grid box of the nitrocellulose membrane and let it dry (either by suction or in incubator).
4. Place the dried membrane on filter paper wetted in denaturation solution for 5 min, in neutralization solution for 1 min, and fixation solution for 20 min.
5. Air dry the membrane and store wrapped in plastic in dark until further use.

3.8.2.2. Prehybridization and Hybridization

1. Incubate the membrane in hybridization buffer at required hybridization temperature (depending upon the melting temperature of oligo probes) for 30 min [Hybridization temperature: $T_{hyb} = T_m - (20–25°C)$].
2. Add heat denatured DIG-labeled probe (1–10 pmol of labeled oligonucleotides per mL of hybridization solution) to the membrane and incubate overnight at the required hybridization temperature.

3.8.2.3. Washing

1. Discard the hybridization solution.
2. Wash two times for 15 min in wash buffer 1 at room temperature.
3. Wash two times for 15 min at hybridization temperature with wash buffer 2.

3.8.2.4. Colorimetric Detection of DIG-Labeled Hybridized Probe

1. Incubate the membrane in blocking solution for 30 min at room temperature.
2. Discard the blocking solution and incubate the membrane in antibody solution for 30 min at room temperature.
3. Discard the antibody solution and remove the surplus antibody conjugates by washing the membrane 2× 15 min with DIG1 buffer.
4. Equilibrate the membrane in detection buffer for 2 min.
5. Add 200 μl of NBT/BCIP stock solution in 10 mL of detection buffer and incubate the membrane in this solution at room temperature in the dark.

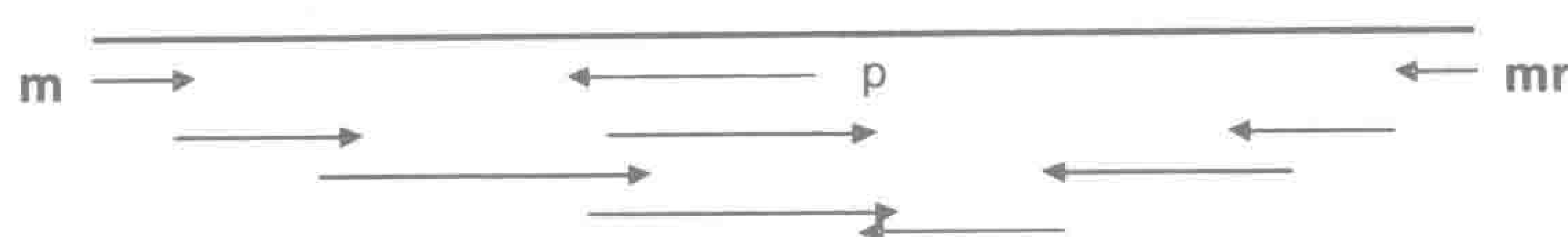


Fig. 5. Sequencing strategy: The target domain that is at an unknown position in the insert region was sequenced using the microarray probe-based primer (p), and both ends were sequenced using M13 forward and reverse primers (m and mr). Further sequencing was carried out using primer walking.

6. After the clear spots become visible, discard the detection buffer, and wash the membrane with H_2O and dry.

After the identification of a positive plasmid clone from the group of plasmid clones by dot blot hybridization, the positive clones can be hybridized to the microarray slides for further confirmation.

3.8.3. Sequencing the Positive Clone

Sequencing of the target gene can be done by primer walking. Two types of primer walking approaches can be applied (Fig. 5). The first approach is to use the universal primers from both ends of the plasmid-insert to target the insert. Sequence information obtained from this initial reaction can be used further to design subsequent primers for further sequencing. In this way, the complete insert sequence or the required gene sequence is obtained. The second approach makes use of the microarray probe-based sequencing primer, which was designed on the basis of the nucleotide sequences of microarray probe that produced the positive signal during the hybridization. Again, a set of sequencing primers can be designed using the sequence information of the positive microarray probe and can be used to attempt sequencing the plasmid sub-clone directly. In principle, the microarray probe-based primers can also be used to sequence the cosmid clone directly.

4. Notes

1. Cloning of the LOV domain from the metagenome-derived novel blue-light photoreceptor:

A 411 bp region encoding for BL receptor LOV-domain was amplified from the cosmid clone using CAGGGACCCGGTAT-GTCGGGCACCCAG (forward) and GGCACCAGAGCGT TCAGCTCCTCGCTCAC (reverse) primers.

As underlined, a 12 bp 5′-end extension was added to the forward primer and a 14 bp 5′-end extension to the reverse primer, generating the vector-specific complementary ends for ligation-independent cloning to a linearized pET52 3C/

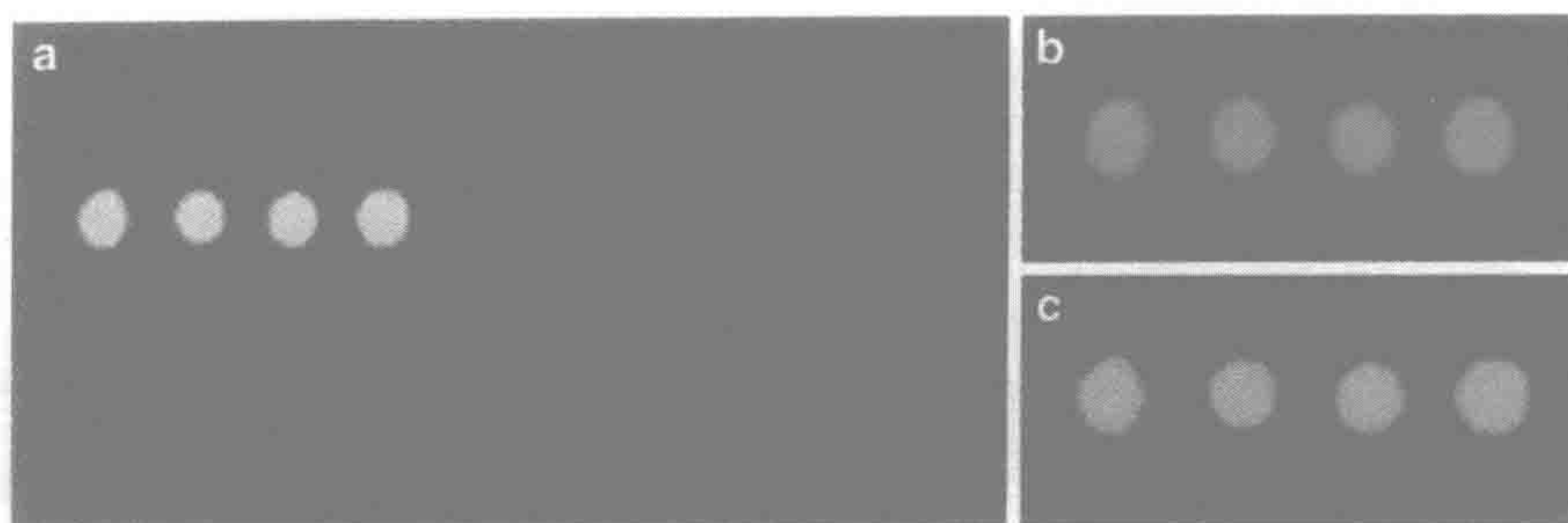


Fig. 6. (**a**) A *bright yellow* image was obtained as the ratio 635/532. (**b**), (**c**) *red* (**b**) and *green* (**c**) images were obtained in single wavelength scans using 635 and 532 nm laser separately.

LIC vector (Novagen). The PCR product obtained was cloned into pET52 3C/LIC cloning vector using the 3C/LIC cloning kit (Novagen). *E. coli* BL21 (DE3) RIL (Stratagene, USA) was used as host for protein expression (0.5 mM IPTG induction for 5 h or overnight). The recombinant protein was purified using Ni^{2+}-IDA resin, and its spectral properties were observed in UV-absorption and Fluorescence spectrometer (for details see (3)).

2. The work described here made use of a cosmid library deposited into 96-well plates. Each of the plate position had been assigned unambiguously to a cosmid colony, picked from a Petri dish. When pooling strategies are applied, only aliquots of the contents of each well are pooled, allowing at a later step of the procedure to reassign a positive clone to a position in the microplate.
3. Positive hybridization images: Image obtained after hybridization of cy5 and cy3 labeled clones generated by restriction digestion of BamHI and SalI restriction enzyme (see Fig. 6). Both signals are overlaid over each other, yielding the yellow color. Here, the BamHI generated clone was labeled with cy5, and the SalI generated clone was labeled with cy3. Both clones were detected first by dot blot hybridization from few clones and were subjected to microarray hybridization for the confirmation.

Acknowledgements

This work has been supported by the Helmholtz Society (VIBS-VH-VI-157) and by a MPG-Ph.D. grant to G.P.

References

1. He, Z., Gentry, T.J., Schadt, C.W., Wu, L., Liebich, J., Chong, S.C., et al. (2007) GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J* 1, 67–77.
2. Yergeau, E., Kang, S., He, Z., Zhou, J., and Kowalchuk, G.A. (2007) Functional microarray analysis of nitrogen and carbon cycling genes across an Antarctic latitudinal transect. *ISME J* 1, 163–179.
3. Pathak, G.P., Ehrenreich, A., Losi, A., Streit, W.R., and Gärtner, W. (2009) Novel blue light-sensitive proteins from a metagenomic approach. *Environ Microbiol* 11, 2388–2399.
4. Christie, J.M., Reymond, P., Powell, G.K., Bernasconi, P., Raibekas, A.A., Liscum, E., et al. (1998) *Arabidopsis* NPH1: a flavoprotein with the properties of a photoreceptor for phototropism. *Science* 282, 1698–1701.
5. Losi, A. (2004) The bacterial counterparts of plant phototropins. *Photochem Photobiol Sci* 3, 566–574.
6. Losi, A. and Gärtner, W. (2008) Bacterial bilin- and flavin-binding photoreceptors. *Photochem Photobiol Sci* 7, 1168–1178.
7. Rhee, S.K., Liu, X., Wu, L., Chong, S.C., Wan, X., and Zhou, J. (2004) Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Appl Environ Microbiol* 70, 4303–4317.
8. West, A.H. and Stock, A.M. (2001) Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem Sci* 26, 369–376.
9. Losi, A., Polverini, E., Quest, B., and Gärtner, W. (2002) First evidence for phototropin-related blue-light receptors in prokaryotes. *Biophys J* 82, 2627–2634.
10. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor, New York, USA.

Chapter 22

Application of DNA Microarray for Screening Metagenome Library Clones

Soo-Je Park, Jong-Chan Chae, and Sung-Keun Rhee

Abstract

Sequence-based screening tools of a metagenome library can expedite metagenome researches considering tremendous metagenome diversities. Several critical disadvantages of activity-based screening of metagenome libraries could be overcome by sequence-based screening approaches. DNA microarray technology widely used for monitoring environmental genes can be employed for screening environmental fosmid and BAC clones harboring target genes due to its high throughput nature. DNAs of fosmid clones are extracted and spotted on a glass slide and fluorescence-labeled probes are hybridized to the microarray. Specific hybridization signals can be obtained only for the fosmid clones that contain the target gene with high sensitivity (10 ng/μL of fosmid clone DNA) and quantitativeness.

Key words: DNA microarray, Fosmid clone library, High-throughput screening, Sequence based, Hybridization, Fluorescent probes, Metagenome microarray

1. Introduction

Metagenomic techniques are used in the analysis of the complex genomes representing complex microbial communities (1, 2). One general approach of metagenomics begins with the preparation of clone libraries that contains large (more than 30 kb) or small (less than 10 kb) inserts directly derived from environmental samples. Novel biocatalysts and metabolites have been obtained from metagenome libraries (3–5). In addition, when clones containing phylogenetic genes, such as the 16S rRNA gene, are retrieved, the DNA sequence information surrounding these genes provides access to the genomes of uncultivated microorganisms and can yield insight into the physiology of such microorganisms (6, 7).

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668,
DOI 10.1007/978-1-60761-823-2_22,

A critical step of metagenome analysis is to screen for clones that contain target genes among a large number of clones. There are two different approaches to screen target gene-containing clones: activity-based and sequence-based screening. In case of activity-based screening, several hundred thousand clones may need to be analyzed in a single screen in order to detect a few functionally active clones (4, 8, 9). This is mainly due to lack of efficient expression of the metagenome-derived genes in the host strains employed. In general, labor-intensive analyses of individual clones or pools of clones within the library are often required for both activity-based and sequence-based screening procedures. Due to the advantage of microarray technology to achieve a high throughput, it has attracted tremendous interest among environmental microbiologists and various types of DNA microarray have been applied to study the microbial diversity of various environments (10–14).

Here, we show a novel format of microarray in which a fosmid library obtained from environmental samples was arrayed on a glass slide. This format of microarray may offer an effective tool for identifying clones from metagenome libraries rapidly and be alternative or complementary to current metagenome screening technologies. The chapter provides a summary of basic techniques describing a sequence-based metagenomic library screening using the microarray.

2. Materials

2.1. Extraction and Purification of Metagenomic DNA

1. Denaturing solution: 4 M Guanidine thiocyanate, 10 mM Tris–HCl, 1 mM EDTA, 0.5% (w/v) sodium lauroyl sarcosinate (Sarkosyl), and 100 mM β-mercaptoethanol (this solution is stable for 1 week).
2. Extraction buffer: 0.1 M Phosphate pH 7.0, 0.1 M ethylenediaminetetraacetic acid (EDTA) pH 8.0, 0.1 M Tris–HCl pH 7.0, 1.5 M NaCl, and 1% (w/v) cetyltrimethylammonium bromide (CTAB). Store at room temperature.
3. Isoamyl chloroform (24:1 chloroform:isoamyl alcohol; CI) (this is toxic and thus care should be taken to prevent exposure to skin, etc.). Store at 4°C.
4. 20% (w/v) Sodium dodecyl sulfate (SDS), isopropanol (absolute), and distilled water (DW).
5. Liquid nitrogen and sand (autoclaved).
6. GELase (Epicentre, Madison, WI).
7. Cesium chloride (CsCl; Sigma-Aldrich, St. Louis, MO).

8. 10 mg/mL Ethidium bromide (EtBr; Bioneer, Korea) (this is a strong mutagen and thus wearing gloves is required. Avoid exposure to direct light.).
9. SeaPlaque® GTG® Agarose (low-melting-point agarose; Cambrex, Rockland, ME) (see Note 1).
10. TAE buffer (1×): 40 mM Tris, 20 mM acetic acid, and 1 mM EDTA. Store at room temperature.

2.2. Construction of Fosmid Library

1. CopyControl™ Fosmid Library Production Kit and T7 control DNA (Epicentre, Madison, WI).
2. Luria–Bertani (LB) and chloramphenicol stock solution (12.5 μg/mL).
3. SeaPlaque® GTG® Agarose (low-melting-point agarose; Cambrex, Rockland, ME).
4. Phage dilution buffer: 10 mM Tris–HCl pH 8.3, 100 mM NaCl, and 10 mM $MgCl_2$. Store at room temperature.
5. TE buffer: 10 mM Tris–HCl pH 7.5 and 1 mM EDTA. Store at room temperature.
6. 3 M Sodium acetate pH 7.0, 1 M $MgSO_4$, chloroform, and 100% ethanol (absolute).
7. Phenol–chloroform–isoamylalcohol (PCI; 25:24:1) (this is toxic and so care should be taken to prevent exposure to skin, etc.). Store at 4°C temperature.
8. Deep 96-well plate (Nalgene, Rochester, NY).
9. Perfectprep Plasmid 96 Vac DB kit (Eppendorf, Germany).
10. CopyControl™ Induction solution (Epicentre, Madison, WI).

2.3. Construction of Microarray

1. Arrayer (Micro Grid II Compact; Genomic Solutions, Hudson, NH).
2. 384-well microplate (U-shape, 384 microarray plate; Genetix, UK).
3. 40% (w/v) Dimethyl sulphoxide (DMSO; Sigma-Aldrich, St. Louis, MO).
4. Amine-coated glass slide (Nuricell, Korea).
5. 95% Ethanol, distilled water (DW), and 0.1% (w/v) SDS.
6. Positive control DNA: Fosmid clones containing the target gene.
7. Glass or metallic microscope slide rack for postprocessing and microscope slide storage boxes.
8. UV Stratalinker 1800 (Stratagene, La Jolla, CA).
9. Quant-iT™ PicoGreen® dsDNA Reagent and Kits (Invitrogen, Carlsbad, CA).

2.4. Fluorescent Labeling of Probe

1. Target gene primer pairs (in this protocol, bacterial 16S rRNA gene: 338F and 517R, 27F, and 1492R).
2. QIAquick PCR purification kit (Qiagen, Germany).
3. BioPrime DNA Labeling kit (Invitrogen, Carlsbad, CA).
4. Cy5 dUTP (Amersham Pharmacia Biotech, Piscataway, NJ).
5. Labeling reaction solution containing 5 mM (each) dATP, dTTP, and dGTP; 2.5 mM dCTP (New England Biolabs, Beverly, MA); 1 mM Cy5 dUTP (Amersham Pharmacia Biotech, Piscataway, NJ).
6. Sodium acetate 40 U of Klenow fragment (Invitrogen, Carlsbad, CA).

2.5. Microarray Hybridization

1. Hybridization solution: 8.75 μL of formamide (50%, v/v), 3× SSC (1× SSC contained 150 mM NaCl and 15 mM trisodium citrate), 1.25 μg of unlabeled herring sperm DNA (Promega, Madison, WI), and 0.3% (w/v) SDS in a total volume of 13.5 μL.
2. Coverslip (10 mm × 15 mm; Sigma-Aldrich, St. Louis, MO).
3. Hybridization chamber (Corning Inc., Corning, NY).
4. Washing solution 1: 1× SSC and 0.2% (w/v) SDS.
5. Washing solution 2: 0.1× SSC and 0.2% (w/v) SDS.
6. Washing solution 3: 0.1× SSC.

2.6. Image Processing and Data Analysis

1. ScanArray 4000 Microarray Analysis system (Perkin-Elmer, Wellesley, MA).
2. GenePix version 6.0 software (Molecular devices Co., Downingtown, PA).
3. Excel 2003 (Microsoft) and Sigmaplot 8.0 (Jandel Scientific, San Rafael, CA).

3. Methods

All the protocols of this instruction are described in three parts (six subheadings). In the first part (Subheadings 3.1–3.3), we provide an efficient protocol for the construction of the fosmid clone library from sediment samples. Efficient extraction of clean and large sized metagenome DNA from samples with high amounts of enzyme inhibitors such as humic substances is the most challenging step for successful construction of metagenome libraries (15–17). In addition, the protocol for the preparation of fosmid DNA from fosmid clone library based on a 96-well plate culture system is explained.

In the second part (Subheading 3.4), the microarray fabrication protocol using the prepared fosmid DNAs is detailed. This protocol is slightly modified from the typical procedure of spotting any nucleic acids on the glass slide using an arrayer and postprocessing of the slides (10, 12).

Finally, the last part (Subheadings 3.5–3.6) describes the protocol for the preparation of the fluorescent probe and hybridization of the probe onto the fosmid DNAs-spotted microarray slides. To avoid confusion in the terminology used in this protocol, the term "probe" is defined as the fluorescence-labeled DNA used for hybridization, "library pool" is defined as the fosmid clones arrayed (spotted) on a glass slide, and "target" is defined to be the clone of the library pool that contains the specific gene complementary to the probe. The overall scheme of these protocols for metagenome microarray is shown in Fig. 1.

3.1. Extraction of Metagenomic DNA from Marine Sediments

1. Extract metagenomic DNA from environmental samples using the modified protocol (17, 18). Add 2 g of sediment or soil sample and 1 g of autoclaved sand in a mortar. Immediately add 1 mL of denaturing solution. Grind to mix soil with denaturing solution.

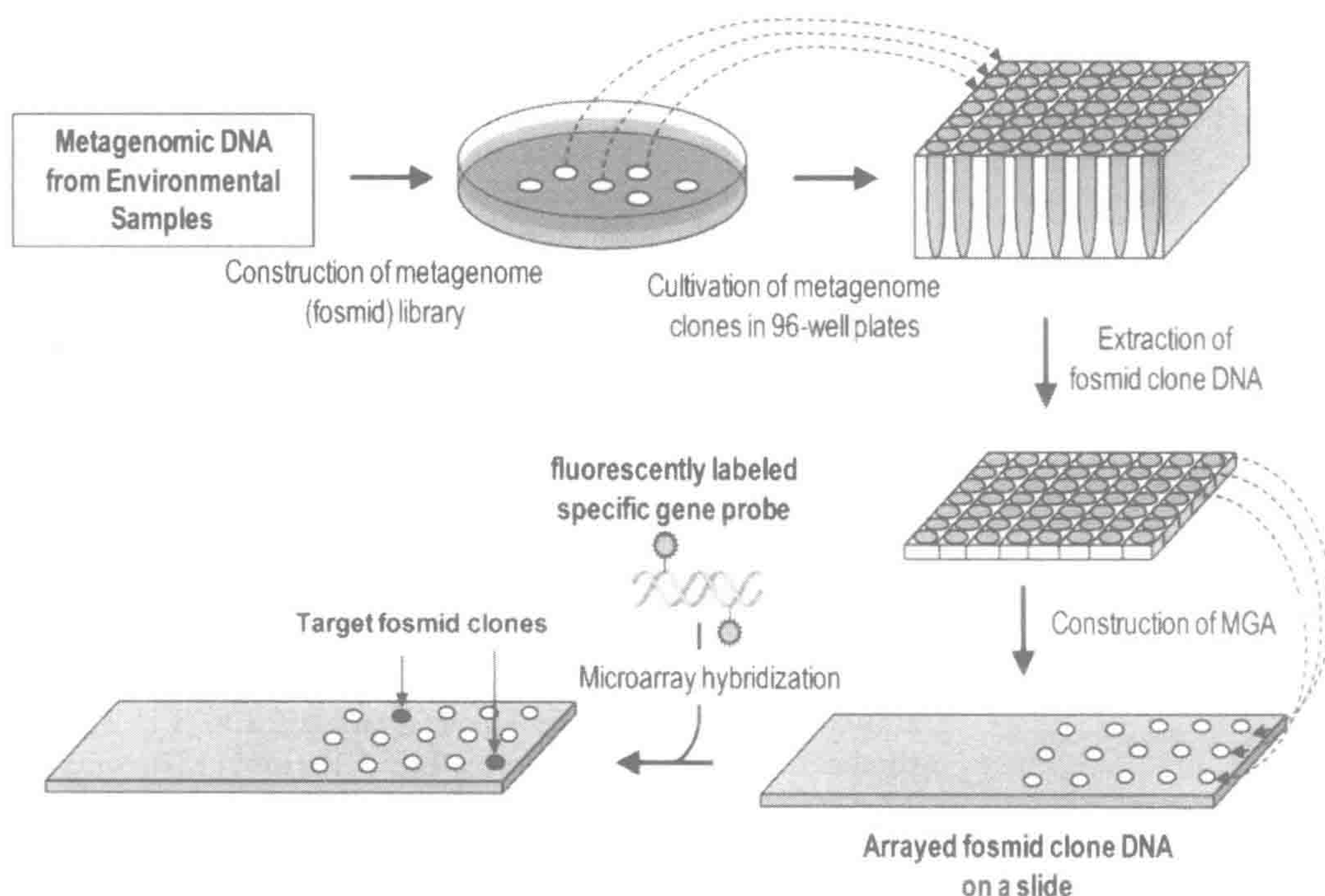


Fig. 1. Schematic diagram of the construction of the metagenome microarray and screening of target clones using the microarray [reproduced from (16) with permission from Wiley-Blackwell].

2. Immediately flood the mixture with liquid nitrogen in a mortar. Grind the sample (avoid thawing the sample). Repeat this freezing and grinding process three times. Transfer sample to a 50-mL polypropylene tube and add 9-mL extraction buffer. Mix gently, incubate at 60°C for 2–3 min and then add 1 mL 20% (w/v) SDS. Gently mix and incubate at 60°C for additional 15 min.
3. Centrifuge at 3,820 × *g* for 10 min at 25°C in a swing bucket rotor and transfer the supernatant into a new 50-mL polypropylene tube. Add an equal volume of CI. After mixing gently, centrifuge at 3,820 × *g* for 20 min at 25°C in a swing bucket rotor.
4. Collect the aqueous upper layer (the upper layer contains the DNA) and add 0.6 volume of isopropanol for precipitation of DNA. Mix gently. Incubate for 30 min. Centrifuge samples at 23,890 × *g* at 25°C for 20 min and obtain precipitated pellet. After air-drying dissolve the DNA pellet with 200 μL DW.
5. The DNA pellets are purified using CsCl gradient centrifugation. Add dissolved DNA to 300 μL DW containing 0.5 g CsCl and 20 μL EtBr gently mix and ultracentrifuge at 500,000 × *g* at 4°C for 12 h and extract DNA bands under UV light. Recover DNA by ethanol precipitation.
6. The extracted DNAs are further purified and fractionated using gel electrophoresis. Separate DNA fragments of >40 kb on the 0.5% (w/v) low-melting-point agarose gel with 1× TAE by electrophoresis at 50 V for 2 h. Excise the gel with an upper genomic DNA band (>40 kb) and digest with GELase for 2 h (see Note 2). Recover DNA by ethanol precipitation and then check the shearing of DNA using gel electrophoresis (see Note 3).

3.2. Library Construction and Preparation of Fosmid DNA

1. The fosmid library is constructed using a CopyControl™ Fosmid Library Production Kit according to the manufacturer's protocol.
2. Purified DNA (0.5 μg) is treated with End-repair Enzyme mix to generate blunt ends and then ligated into the fosmid vector pCC1FOS.
3. After in vitro packaging into lambda phages and then transfection into the *Escherichia coli* EPI300-T1^R (*E. coli*), the bacterial cells are plated on Luria–Bertani (LB) containing 12.5 μg/mL chloramphenicol.
4. The plates are incubated at 37°C for 24 h prior to the selection of colonies. Transfected *E. coli* colonies are transferred to 96-well plates containing 12.5 μg/mL chloramphenicol-containing LB liquid medium with induction solution and are incubated at 37°C for 24 h.

5. The plates are replicated into deep 96-well plates with 1.5 mL of LB in each well and incubated in a shaking incubator at 180 rpm and 37°C in the presence of chloramphenicol and an inducer which is supplied by the manufacturer (Epicentre, Madison, WI).
6. Transfer 100 μL of culture to a regular 96-well plate and add 100 μL of 30% (w/v) glycerol. Mix gently and store the plates at −70°C as a library.
7. Extra cells are harvested using an appropriate centrifuge for plates (850 × g for 30 min at 4°C) and the fosmid DNA is extracted using a Perfectprep Plasmid 96 Vac DB kit according to the manufacturer's protocol.
8. The fosmid DNA samples are resuspended in DW as a final concentration of ca. 40 ng/μL.

3.3. Microarray Construction and Postprocessing

1. Ten microliters of each fosmid DNA in the 96-well plate are transferred to a 384-well microplate (see Note 4). Comparative study between intact fosmid DNA and restriction enzyme-digested fosmid DNA showed no differences in hybridization signal. Thus, extracted fosmid DNA is directly used for printing without any pretreatment.
2. In the 384-well microplate, the DNA samples are diluted with equal volume of 40% (v/v) DMSO (see Note 4) and then gently mix using multichannel pipette.
3. Use clones containing target gene as a positive control and position marker. A clone with any nontarget insert is used as a negative control.
4. Array the fosmid DNA samples and control DNAs from the 384-well microplate onto amine-coated glass slide at 55–58% relative humidity. The distance between dots is 400 μm. Each fosmid DNA is printed in four different positions of the glass slide as replicates. This is performed with Micro Grid II Compact Arrayer. These instructions are easily adaptable to other arrayers.
5. After printing, the slide was cross-linked by UV irradiation with UV Stratalinker 1800. The slides are exposed to 80 mJ of UV irradiation and washed at room temperature with 0.1% (w/v) SDS for 4 min, followed by washing with water for 2 min. To denature template DNAs, the slide is boiled for 10 min and suddenly placed in cold water.
6. The slides are dried by centrifugation at 950 × g for 5 min and stored in a black slide box at room temperature (see Note 5).

3.4. Fluorescent Labeling of Probe

1. In this protocol, we select 16S rRNA gene as a target. Amplify 16S rRNA gene sequences from the metagenome, which is

used for the construction of fosmid clone library. PCR products of two different lengths (200 and 1,500 bp) are amplified with two different primer sets, respectively: 338F and 518R (19); 27F and 1492R (20, 21). The PCR amplification conditions are as follows: 95°C for 5 min for 1 cycle; 94°C for 30 min, 55°C for 30 s, and 72°C for 60 s for 30 cycles; then 72°C for 5 min for 1 cycle. The PCR product is purified using a QIAquick PCR purification kit (see Note 6).

2. The labeling of 100 or 500 ng PCR product is performed with the BioPrime DNA Labeling kit. Probe DNA (each concentration) is mixed with 15 μg of random octamer, denatured by boiling for 2 min, and immediately chilled on ice. The denatured probe DNA is then mixed with 15 μL of the labeling reaction solution, and the reaction mixture is incubated at 37°C for 3 h (see Note 7).
3. The labeled probe is purified using a QIAquick PCR purification kit, dried completely in a vacuum evaporator for 1 h, resuspended in 4.5 μL of DW, and then stored at –20°C (see Note 7).

3.5. Microarray Hybridization

1. The fluorescence-labeled probe (4.5 μL) is mixed with hybridization solution (13 μL). Heat a part of the hybridization solution mixture (a reduced volume: 7.5 μL) for 5 min for probe denaturation at 95°C using a PCR machine (see Note 8). Deposit the denatured probe directly onto the slides and cover with the coverslip. The slide should be prewarmed at 60°C for 5 min (see Note 9).
2. Dispense 15 μL of 3× SSC into the hybridization wells on either side of a hybridization chamber to prevent evaporation of hybridization solution mixture. Plunge the hybridization chamber immediately into the 50°C water bath (see Note 10). The time that the slide remained at room temperature should be minimized to prevent cross-hybridization.
3. Take out each microarray slide from the hybridization chamber. Remove the coverslip immediately in wash solution 1 [1× SSC and 0.2% (w/v) SDS] and wash the slide at the same time. Wash the slide successively using wash solution 2 [0.1× SSC and 0.2% (w/v) SDS] and wash solution 3 (0.1× SSC) for 5 min each at ambient temperature prior to being dried.
4. The slides are dried using centrifugation. Put the slide in the conical tube and centrifuge at 420 ×*g* for 5 min. Keep the slide in the black slide box at room temperature before scanning.

3.6. Image Processing and Data Analysis

1. The microarray slides are scanned by using a ScanArray 4000 Microarray Analysis system at a resolution of 10 μm. The laser power and photomultiplier tube (PMT) gain are adjusted to

avoid saturation of target spots, and the scanned images are saved as 16-bit TIFF files (see Note 11).

2. Each spot is quantified using GenePix version 6.0 software. For gridding, positive control spots are necessary for positioning as landmarks. An example of the scanned image with circular grids of the software for each fosmid clone is shown in Fig. 2a. In cases of strong signal, the spot can be seen with the naked eye with contrast adjustment (2F8 and 4D5 of Fig. 2a). The signal-to-noise ratio (SNR) is used as the hybridization signal intensity. The SNR is calculated with the following formula (22) as a basis: SNR = (signal intensity – background)/standard deviation of background. In the formula, the background measurement refers to the local spot background intensity, and the standard deviation of the background is calculated by the GenePix version 6.0 software. After measurement of signal intensity, the statistics are performed with Excel 2003 and Sigmaplot 8.0.
3. The SNRs from four replicate data sets are averaged to represent the SNR for a particular fosmid clone. An example of the calculated signal intensities (SNR) for the fosmid clones containing 16S rRNA gene is shown in Fig. 2b.

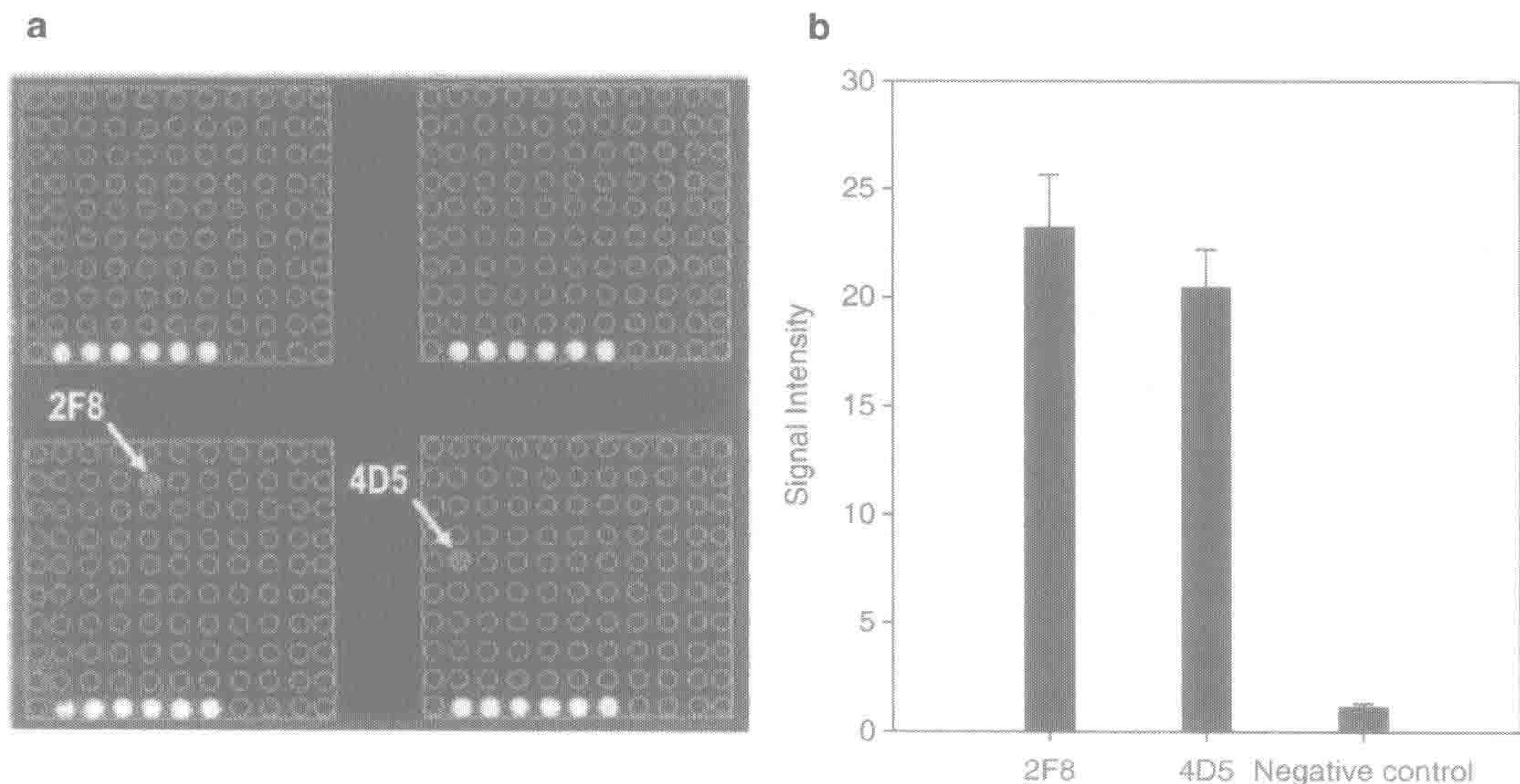


Fig. 2. Fluorescence images showing hybridization of the metagenome microarray constructed using a fosmid library from marine sediment. (**a**) A circular grid was overlapped onto the fluorescent image using GenePix version 6.0 software to aid identification of arrayed spots. *Arrows* indicate the position of spots of 16S rRNA gene (target)-containing clones. The PCR product of the 16S rRNA gene (50 ng/mL) amplified from the genomic DNA that was obtained from the marine sediment was arrayed on the *last row* of *each panel* of the slide (*six spots*) and used as a positive control, which was saturated in this hybridization and scanning condition and shown by a *white color*. (**b**) The signal intensities of the candidate fosmid clone spots and negative spots were calculated and compared [reproduced from (16) with permission from Wiley-Blackwell].

4. Notes

1. We suggest using a performance certified (high gel strength) low-melting agarose for high molecular weight DNA separation in preparative DNA electrophoresis. For easy handling of low-melting agarose, gel strength should be confirmed before use.
2. Do not expose the sample DNA to UV irradiation. Even short UV exposure can decrease the efficiency of cloning and ligation by two to three orders of magnitude or more.
3. Frequently genomic DNA is sufficiently sheared as a result of the purification process, additional shearing is not necessary. Be sure to cut the gel slice so that the DNA recovered is ≥25 kb. Cloning DNA smaller than about 25 kb may result in unwanted chimeric clones.
4. The minimum volume of spotting material in the 384-well plate is ca. 10 μL. To reduce the minimum volume to this level, we suggest using 384-well plates with conical or round shape bottom. About 15 μL of solution is more than enough since one dipping of the spotting pin takes very small volume of spotting solution (ca. 0.25 μL) depending on the pins and can be used for spotting several dozens of slides. The plate should be sealed and kept at –80°C to avoid evaporation and opened only during spotting.
5. The slides can be stored at room temperature for a couple of months without deterioration. For longer storage more than a year, keep the slides in conical tube and store at –80°C.
6. Shorter PCR products (≤200 bp) and labeled long oligonucleotide (≤70 bp) can be used as probes for hybridization depending on the research purpose.
7. During this step, avoid exposure to direct light.
8. To save the slide, several probes could be used simultaneously in a hybridization reaction. In this case, each probe should be labeled using different dye-dCTP. The scanner should have corresponding filter to each dye.
9. Extra hybridization mixture solution can be kept at –20°C but should be used in a week.
10. It is important to keep the hybridization solution (and slide) over the hybridization temperature (50°C) during the whole procedure until washed to prevent potential cross-hybridization. This is especially important when single-stranded nucleic acids are used as the probe (or target) (12).
11. The adjustment of laser power and PMT gain does not affect SNR value unless the signal is saturated although the intensity of spot looks different.

Acknowledgments

This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant R01-2007-000-20806-0 and program 2007-04269 (System development for application of genomic sequence information) funded by the Korean Government (MEST).

References

1. Kowalchuk, G.A., Speksnijder, A.G., Zhang, K., Goodman, R.M., and van Veen, J.A. (2007) Finding the needles in the metagenome haystack. *Microb Ecol* 53, 475–485.
2. Schmeisser, C., Steele, H., and Streit, W.R. (2007) Metagenomics, biotechnology with non-culturable microbes. *Appl Microbiol Biotechnol* 75, 955–962.
3. Williamson, L.L., Borlee, B.R., Schloss, P.D., Guan, C., Allen, H. K., and Handelsman, J. (2005) Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. *Appl Environ Microbiol* 71, 6335–6344.
4. van Hellemond, E.W., Janssen, D.B., and Fraaije, M.W. (2007) Discovery of a novel styrene monooxygenase originating from the metagenome. *Appl Environ Microbiol* 73, 5832–5839.
5. Kim, B.S., Kim, S.Y., Park, J., Park, W., Hwang, K.Y., Yoon, Y. J., et al. (2007) Sequence-based screening for self-sufficient P450 monooxygenase from a metagenome library. *J Appl Microbiol* 102, 1392–1400.
6. Hallam, S.J., Mincer, T.J., Schleper, C., Preston, C.M., Roberts, K., Richardson, P.M., et al. (2006) Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. *PLoS Biol* 4, e95.
7. Hallam, S.J., Konstantinidis, K.T., Putnam, N., Schleper, C., Watanabe, Y., Sugahara, J., et al. (2006) Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc Natl Acad Sci USA* 103, 18296–18301.
8. Majernik, A., Gottschalk, G., and Daniel, R. (2001) Screening of environmental DNA libraries for the presence of genes conferring Na(+)(Li(+))/H(+) antiporter activity on *Escherichia coli*: characterization of the recovered genes and the corresponding gene products. *J Bacteriol* 183, 6645–6653.
9. Henne, A., Schmitz, R.A., Bomeke, M., Gottschalk, G., and Daniel, R. (2000) Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl Environ Microbiol* 66, 3113–3116.
10. Bae, J.W., Rhee, S.K., Park, J.R., Chung, W.H., Nam, Y.D., Lee, I., et al. (2005) Development and evaluation of genome-probing microarrays for monitoring lactic acid bacteria. *Appl Environ Microbiol* 71, 8825–8835.
11. Bodrossy, L., Stralis-Pavese, N., Murrell, J.C., Radajewski, S., Weilharter, A., and Sessitsch, A. (2003) Development and validation of a diagnostic microbial microarray for methanotrophs. *Environ Microbiol* 5, 566–582.
12. Rhee, S.K., Liu, X., Wu, L., Chong, S.C., Wan, X., and Zhou, J. (2004) Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Appl Environ Microbiol* 70, 4303–4317.
13. Ward, B.B., Eveillard, D., Kirshtein, J.D., Nelson, J.D., Voytek, M.A., and Jackson, G.A. (2007) Ammonia-oxidizing bacterial community composition in estuarine and oceanic environments assessed using a functional gene microarray. *Environ Microbiol* 9, 2522–2538.
14. Wu, L., Thompson, D.K., Liu, X., Fields, M.W., Bagwell, C.E., Tiedje, J.M., et al. (2004) Development and evaluation of microarray-based whole-genome hybridization for detection of microorganisms within the context of environmental applications. *Environ Sci Technol* 38, 6775–6782.
15. Hallam, S.J., Putnam, N., Preston, C.M., Detter, J.C., Rokhsar, D., Richardson, P.M., et al. (2004) Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* 305, 1457–1462.
16. Park, S.J., Kang, C.H., Chae, J.C., and Rhee, S.K. (2008) Metagenome microarray for screening of fosmid clones containing specific genes. *FEMS Microbiol Lett* 284, 28–34.
17. Quaiser, A., Ochsenreiter, T., Klenk, H.P., Kletzin, A., Treusch, A.H., Meurer, G., et al.

(2002) First insight into the genome of an uncultivated crenarchaeote from soil. *Environ Microbiol* 4, 603–611.

18. Zhou, J., Bruns, M.A., and Tiedje, J.M. (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbiol* 62, 316–322.
19. Muyzer, G., de Waal, E.C., and Uitterlinden, A.G. (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* 59, 695–700.
20. Park, S.J., Kang, C.H., and Rhee, S.K. (2006) Characterization of the microbial diversity in a Korean solar saltern by 16S rRNA gene analysis. *J Microbiol Biotechnol* 16, 1640–1645.
21. Weisburg, W.G., Barns, S.M., Pelletier, D.A., and Lane, D.J. (1991) 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* 173, 697–703.
22. Verdnick, D., Handran, S., and Pickett, S. (2002) Key considerations for accurate microarray scanning and image analysis. In: Kamberova, G. Ed. DNA Image Analysis: Nuts and Bolts. DNA Press, Salem, MA, pp. 83–98.

Chapter 23

MetaGenomeThreader: A Software Tool for Predicting Genes in DNA-Sequences of Metagenome Projects

David J. Schmitz-Hübsch and Stefan Kurtz

Abstract

We consider a gene finding method that is specifically designed to work on metagenome sequences. The method can handle short metagenome sequences with in-frame stop codons as well as frame shifts. It delivers gene predictions for a set of metagenome sequences, which may be individual reads or a collection of assembled reads sequenced from an environmental sample. The method searches for stretches of DNA that are conserved within the environmental sample. Conserved coding sequences are discriminated from conserved non-coding regions based on their synonymous substitution rate. We describe the program *MetaGenomeThreader* which implements the method and show its application on a synthetic metagenome.

Key words: Metagenome, Gene prediction, Dynamic programming, Combined scoring matrix, Frameshift postprocessing

1. Introduction

Sequences from metagenome projects are often short and they contain in-frame stop codons as well as frame shifts. These constraints make it difficult to reliably predict genes in metagenome sequences. Krause et al. (1) have developed a novel gene finding method that is specifically designed to work under these constraints. The method improves on previous methods and delivers gene predictions for a set of metagenome sequences, which may be individual reads or a collection of assembled reads sequenced from an environmental sample. The method searches for stretches of DNA that are conserved within the environmental sample. Conserved coding sequences are discriminated from conserved non-coding regions based on their synonymous substitution rate. The larger the ratio of synonymous to non-synonymous substitutions, the

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668, DOI 10.1007/978-1-60761-823-2_23, © Springer Science+Business Media, LLC 2010

more likely the sequence codes for a functional protein. This feature was observed several times before and it was used in previous gene finders, e.g., in (2, 3).

The method of (1) is based on pairwise sequence comparisons of the metagenome sequence and a database of nucleotide sequences, e.g., a database of complete genomes, metagenomes, or known genes. The main idea is to combine the information from the pairwise sequence comparisons such that these are considered simultaneously if they refer to the same region of the metagenome sequence. To discriminate if the pairwise sequence comparisons reveal conserved coding sequences, conserved non-coding regions, or shadows of coding sequences in another reading frame, the method finds the best combination of partial sequence comparisons, taking into account (a) the synonymous substitution rate at each position in the metagenome sequence, (b) the positions of stop codons in the metagenome sequence, and (c) the position of stop codons in the nucleotide database. The use of the best combination for all pairwise sequence comparisons is expected to result in a superior performance compared to methods that rely on simple pairwise sequence alignments only.

In this chapter we will explain the method of (1). While the original description is very compact, we take time to describe the different steps in more detail and fill gaps to ease understanding of the method. At the end of the Methods section, we describe *MetaGenomeThreader*, our own implementation of the method described here. We show how to apply *MetaGenomeThreader* to a synthetic metagenome and discuss the results achieved in the application.

2. Materials

The input for the *MetaGenomeThreader* algorithm are nucleotide sequences from a metagenome project and a large collection of nucleotide sequences, serving as a reference. The reference sequences are used as the database and the metagenome sequences are used as the query for the TblastX-program, a subprogram of the standalone blastall program provided by the National Center for Biotechnology Information (NCBI). TblastX translates the DNA sequences of the query and the database in all six reading frames and outputs significant pairwise alignments of the query sequence and the database sequence on the protein level. These alignments serve as input to the *MetaGenomeThreader* algorithm. The *MetaGenomeThreader* data flow thus consists of the following steps:

1. Collect metagenome sequences.
2. Collect reference sequences (e.g., the nt-database).

3. Apply the formatdb-program from the standalone version of NCBI Blast to the reference sequences.
4. Run TblastX using the reference sequences as database and the metagenome sequences as queries.
5. Apply *MetaGenomeThreader* on the TblastX results obtained in the previous step to obtain predicted coding sequences.

The methods underlying step 5 and some applications of the methods will be explained in more detail in the following sections.

3. Methods

3.1. Basic Notation

For any sequence S of length m, $S[i]$ denotes the character at position i in S, for $1 \leq i \leq m$. For $i \leq j$, $S[i...j]$ denotes the substring of S starting with the character at position i and ending with the character at position j.
Let $\mathcal{F} = \{+1,+2,+3,-1,-2,-3\}$ be the set of reading frames. The positive numbers denote the reading frames on the forward strand, and the negative numbers denote the reading frames on the reverse strand. A codon is a sequence of length 3 over the DNA alphabet. For each sequence S of length m and for each reading frame $f \in \mathcal{F}$,

$$\alpha_f(S,l) := \begin{cases} f + 3(l-1), & \text{if } f > 0, \\ m+1+f-3(l-1), & \text{if } f < 0 \end{cases}$$

is the first position of the lth codon in S. Hence,

$$c_f(S)[l] := \begin{cases} S[\alpha_f(S,l)]S[\alpha_f(S,l)+1]S[\alpha_f(S,l)+2], & \text{if } f > 0, \\ S[\alpha_f(S,l)]S[\alpha_f(S,l)-1]S[\alpha_f(S,l)-2], & \text{if } f < 0 \end{cases}$$

is the lth codon of sequence S in reading frame f. So $c_f(S)$ is the sequence of codons from s in reading frame f. A codon alignment is a sequence of columns, where each column either consists of one codon above another codon, a gap symbol above a codon, or a codon above a gap symbol.

3.2. Phases of the Algorithm

3.2.1. First Phase: Identification of Conserved Sequences

This phase treats each metagenome sequence separately. So consider a metagenome sequence M (e.g., a single 454-read or an assembled contig of 454-reads (4) from a metagenome sequencing project) and a nucleotide database R (e.g., Genbank). In the first phase of the algorithm, local sequence alignments of M and R are computed, i.e., alignments involving continuous parts of M and continuous parts of R (but not necessarily the entire sequences). Thus, R plays the role of a reference sequence,

which is compared to the metagenome sequence to derive information valuable for predicting coding sequences. To make the comparison sensitive, the local alignments are computed on the protein level, i.e., M and R are first translated into all six reading frames before computing significant local alignments on the translated amino acid sequences. As the reading frame in each local amino acid alignment is known, it can uniquely be transformed to an alignment of codons from the original sequence. Thus, the result of the alignment process is a set of high-scoring local codon alignments in specified reading frames. Note that codon alignments align codons and they can contain insertions and deletions, but only of complete codons.

Now consider a codon alignment with two corresponding codons $a_1a_2a_3$ in M and $b_1b_2b_3$ in R. That is, $a_1a_2a_3$ and $b_1b_2b_3$ appear in the same column of the codon alignment. Position $r \in \{1,2,3\}$ contains a synonymous base exchange in codon pair $(a_1a_2a_3, b_1b_2b_3)$ if $a_r \neq b_r$ and the codons translate to the same amino acid. Position $r \in \{1,2,3\}$ has a non-synonymous base exchange in codon pair $(a_1a_2a_3, b_1b_2b_3)$ if $a_r \neq b_r$ and the codons translate to a different amino acid. See Fig. 1 for an example. For each local codon alignment, say A, one counts the number $\sigma(A)$ of positions with synonymous base exchanges and the number $\bar{\sigma}(A)$ of positions with non-synonymous base exchanges. If $\sigma(A)$ is smaller than $\bar{\sigma}(A)$, then this indicates that the aligned part of R is not coding and hence, the local codon alignment is discarded. If $\sigma(A) \geq \bar{\sigma}(A)$, then the local codon alignment is processed further in the second phase, described next.

```
D   C   K   E   I   L   A   D   L   L   L   F
    |   |   |   |   |   |   |   |   |
H   C   K   E   I   L   A   D   L   L   -   -

GAC TGC AAA GAG ATT CTA GCT GAT CTG CTG CTA TTT
CAC TGC AAG GAA ATC CTA GCT GAT CTG CTG --- ---
n         s   s   s

D   C   K   E   I   L   A   D   L   L   L   F
|               |   |   |   |   |   |   |
D   -   -   -   I   L   A   D   L   L   L   I

GAC TGC AAA GAG ATT CTA GCT GAT CTG CTG CTA TTT
GAC --- --- --- ATC CTA GCT GAT CTT CTG CTA ATT
                  s               s         n
```

Fig. 1. Two amino acid alignments and the corresponding codon alignments. Positions with synonymous and non-synonymous base exchanges are marked by s and n, respectively.

3.2.2. Second Phase: Computation of the Combined Scoring Matrix

This phase combines all local codon alignments for the same metagenome sequence M. To do so, for all pairs of codons in the local codon alignments one computes scores that are (for each reading frame) accumulated at the corresponding positions of M, see Fig. 2 for an example.

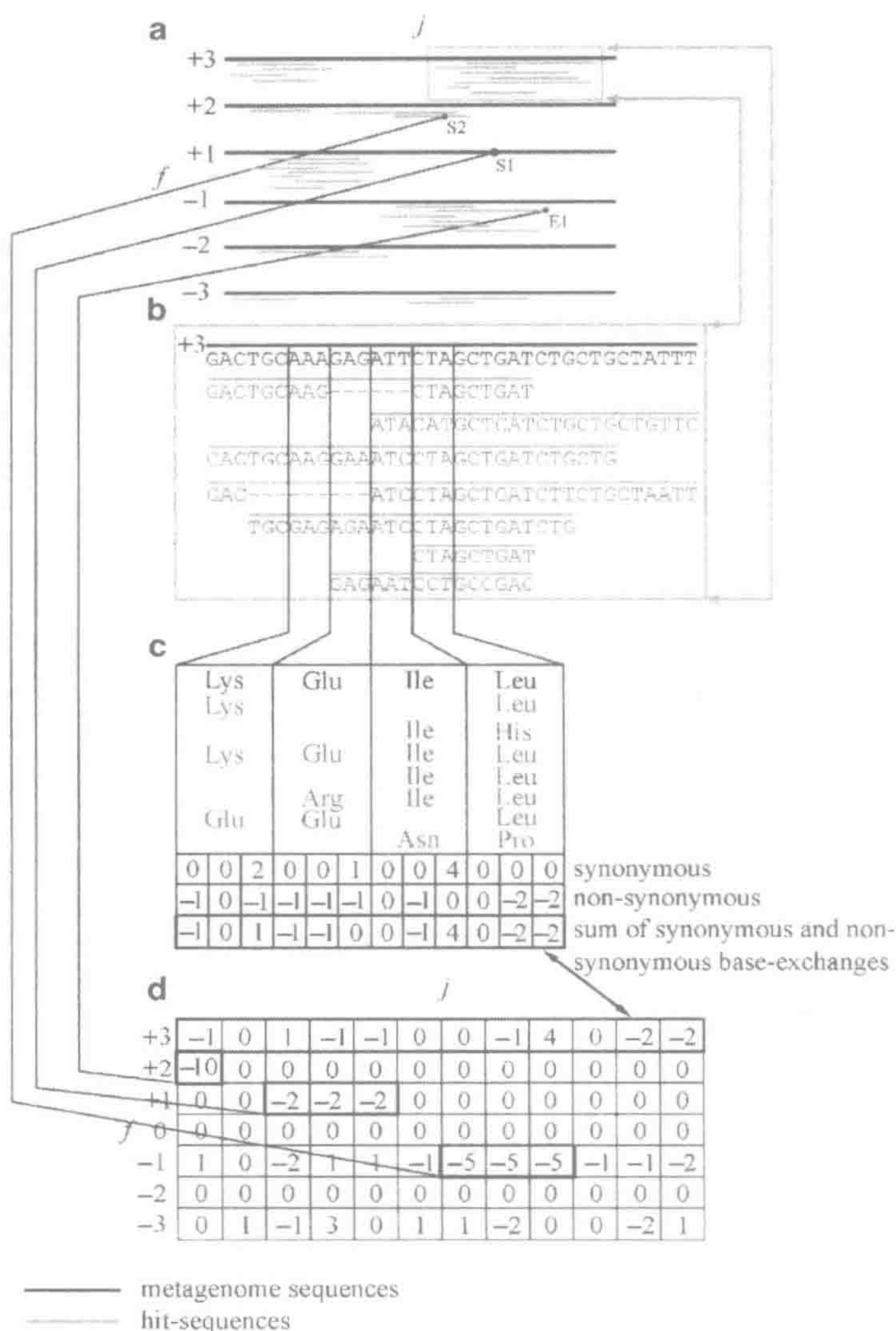


Fig. 2. Calculating the absolute scores of the combined score matrix. (**a**) All six reading frames of a metagenome sequence are shown as a *black line*. Regions of the reference sequence involved in the local codon alignments are shown as *grey lines* below a line depicting the reading frame. (**b**) The metagenome sequence in reading frame + 3 is shown above the regions of the reference sequence (shown in *light grey*) involved in the local codon alignment. (**c**) Synonymous and non-synonymous codon base exchanges lead to different scores which are accumulated at the different positions of the metagenome sequence. (**d**) The accumulated scores for each reading frame and each position are stored in a matrix of seven rows, where the 0-row with zero scores is used to model a non-coding region. Penalties are added for stop codons in the metagenome (case S1) or in the reference (case S2), or when reaching the ends of aligned parts of the reference sequence (case E1).

Let $\mathcal{F}_0 = \mathcal{F} \cup \{0\}$ and C_M be an $\mathcal{F}_0 \times \{1,\ldots,m\}$-table which is initialized to 0 for each $f \in \mathcal{F}_0$ and each $j \in \{1,\ldots,m\}$. Let $\alpha = a_1a_2a_3$ and $\beta = b_1b_2b_3$ be the codons in frame f as defined above. Suppose that codon $a_1a_2a_3$ starts at position j in M. Then for each $r \in \{1,2,3\}$ one adds a score $\gamma(r)$ to $C_M(f, j+r-1)$ where $\gamma(r)$ is defined as follows:

- If position $j+r-1$ corresponds to the last position of R, then this is considered an indicator for gene boundaries. Therefore, one sets $\gamma(r) := -10$ to penalize this case.
- In-frame stop-codons are handled separately for M and R. More precisely, if α is a stop codon and β is not a stop codon, then $\gamma(r) = -2$. If β is a stop codon, then $\gamma(r) = -5$.
- If α and β translate into different amino acids and $a_r \neq b_r$, then $\gamma(r) = -1$. That is, r is one of three possible codon positions with a base exchange that leads to a different translation. This is penalized by a negative score.
- If α and β translate into the same amino acid and $a_r \neq b_r$, then $\gamma(r) = 1$. That is, a silent mutation in the r-th position of the codon is rewarded with a positive score.
- If $a_r = b_r$, then $\gamma(r) = 0$. That is, a position at which there is no base change does not influence the score.

Applying this scoring scheme to all codon pairs in all local codon alignments one obtains a combined score matrix C_M for the metagenome sequence M:

$$\begin{array}{ccccccc}
C_M(+3,1) & \cdots & C_M(+3,j-1) & C_M(+3,j) & C_M(+3,j+1) & \cdots & C_M(+3,m), \\
C_M(+2,1) & \cdots & C_M(+2,j-1) & C_M(+2,j) & C_M(+2,j+1) & \cdots & C_M(+2,m), \\
C_M(+1,1) & \cdots & C_M(+1,j-1) & C_M(+1,j) & C_M(+1,j+1) & \cdots & C_M(+1,m), \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0, \\
C_M(-1,1) & \cdots & C_M(-1,j-1) & C_M(-1,j) & C_M(-1,j+1) & \cdots & C_M(-1,m), \\
C_M(-2,1) & \cdots & C_M(-2,j-1) & C_M(-2,j) & C_M(-2,j+1) & \cdots & C_M(-2,m), \\
C_M(-3,1) & \cdots & C_M(-3,j-1) & C_M(-3,j) & C_M(-3,j+1) & \cdots & C_M(-3,m)
\end{array}$$

Note that the matrix entries $C_M(0, j)$ never change, i.e., they remain 0. So the matrix contains a 0-row which is introduced for technical reasons explained later. To make the score at the different positions of M comparable, one normalizes the matrix entries by dividing the score $C_M(f, j)$ by the number of local codon alignments covering position j. This is done for all reading frames. The resulting matrix is the normalized combined score matrix, denoted by $\overline{C}_M$. The entry $\overline{C}_M(f, j)$ reflects the coding potential of position j in reading frame f. The higher $\overline{C}_M(f, j)$, the higher the coding potential. The next step considers how to predict coding sequences from matrix $\overline{C}_M$.

3.2.3. Third Phase: Prediction of Coding Sequences

The *coding graph* $G(M)$ of M is an edge-labeled graph. The nodes are the pairs (f, j) where $f \in \mathcal{F}_0$ and $j \in \{1,\ldots,m\}$. For each j, $2 \le j \le m$, there is an edge from $(f', j-1)$ to (f, j), denoted by $(f', j-1) \to (f, j)$, if and only if $f' \in \pi(f, j)$,

where

$$\pi(f,j) = \begin{cases} \{z, 0, -z\}, & \text{if } f = 0, \\ \{f, 0, -f\}, & \text{else if } |f| = z, \\ \{f\}, & \text{otherwise,} \end{cases}$$

and $z = (j-1) \bmod 3 + 1. \pi(f,j)$ is the set of precursor frames allowed for position j in frame f. Let q be some negative score value. An edge $(f', j-1) \to (f, j)$ in the graph is labeled by a score $\bar{C}_M(f,j) + \delta(f', f)$, where $\delta(f', f)$ is defined as follows:

- If $f < 0$ and $f' > 0$ or $f > 0$ and $f' < 0$, then $\delta(f', f) = 2q$. That is, changing the direction of the reading frame from forward to reverse or vice versa gives a penalty of $2q$.
- If $f \ne 0$, $f' \ne f$ and either $f, f' > 0$ or $f, f' < 0$, then $\delta(f', f) = q$. That is, changing the reading frame while keeping the reading direction invariant gives a penalty of q.
- If $f = 0$ or $f = f'$, then $\delta(f, f') = 0$. That is, there is no extra penalty in the non-coding state 0 or when the reading frame remains invariant.

Consider a path $(f_1, 1) \to (f_2, 2) \to \cdots \to (f_{m-1}, m) \to (f_m, m)$ in $G(M)$. The total score of the path is the sum of the scores along the edge labels plus the initial score $\bar{C}_M(f_1, 1)$. By projection onto the first component, one obtains a sequence of frames $f_1, f_2, \cdots, f_{m-1}, f_m$, which has the same score as the corresponding path in $G(M)$. The goal is to find a sequence of frames with maximum score. This is obtained by a dynamic programming algorithm, which computes a matrix N such that for each $f \in \mathcal{F}_0$ and each $j \in \{1,\ldots,m\}$, $N(f, j)$ is the maximum score of any path beginning with some $(f', 1)$ and ending in (f, j). N can be computed by the following equation:

$$N(f,j) = \begin{cases} \bar{C}_M(f,j), & \text{if } j = 1, \\ \max\{N(f', j-1) + \bar{C}_M(f,j) + \delta \mid (f', f) \mid f' \in \pi(f,j)\}, & \text{otherwise.} \end{cases}$$

As for each j one has to maximize over a constant number of values, each of which is computed in constant time, matrix N can be evaluated in time proportional to the length m of M. By computing $\max\{N(f,m) \mid f \in \mathcal{F}_0\}$ one obtains the maximum score of the last column of table N. Backtracing from this value gives the maximum scoring path from the first column to the last column of the matrix, and thus the sequence of reading frames of

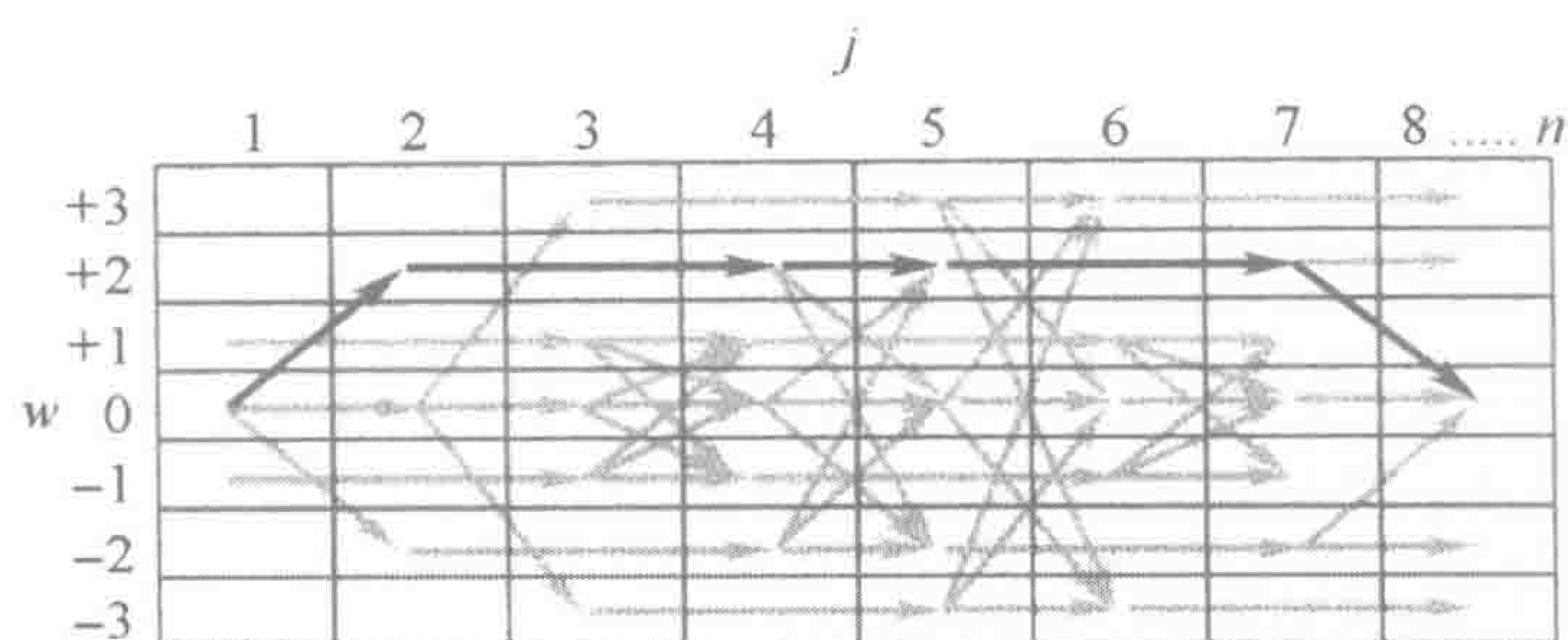


Fig. 3. Predicting coding sequences by calculating the optimal path in matrix *N*.

maximum score. See Fig. 3 for an example. Each maximal subpath of identical reading frames $f \neq 0$ is considered a predicted coding sequence (PCS, for short). The set of all coding sequences is denoted by P^*.

3.2.4. Fourth Phase: Postprocessing the Predicted Coding Sequences

If the reference sequences consists of many short sequences, genes in long metagenome sequences may only partially be covered by codon alignments. This usually leads to predictions consisting of several predicted coding sequences for the target gene. A post-processing step then needs to join the different coding sequences, taking frame shifts into consideration. Frameshifts are due to insertions and deletions of single bases resulting from errors in the sequencing process. The algorithm considers potential sequencing errors by checking the distance between different PCSs (1). Adjacent predictions within the same reading frame are joined if their distance on *M* does not exceed some threshold and the coding sequence does not have an in-frame stop codon, whereas adjacent predictions in different reading frames are joined if their distance on *M* does not exceed some threshold and the PCSs do not have an in-frame stop codon due to the correct reading frame. See Fig. 4 for an example. If the metagenome sequences are short contigs (e.g., 454 reads), the postprocessing phase also makes sense, as it joins non-continuous alignments (gaps in the AA-sequence) and detects frame-shifts of single sequence-positions.

3.3. Implementation of the Methods

The algorithms described in this section were implemented in the software called *MetaGenomeThreader*, see (5). *MetaGenome Threader* is available as part of the open source software Genome Tools software package (6). We use TblastX (7) to compute local amino acid alignments of the metagenome sequence and the reference sequence in all reading frames. From the amino acid alignments *MetaGenomeThreader* computes the corresponding codon alignments in the specific reading frame.

3.4. Application of the Methods

We report some results when applying *MetaGenomeThreader* to a synthetic metagenome from a 454-sequencing project. The use of

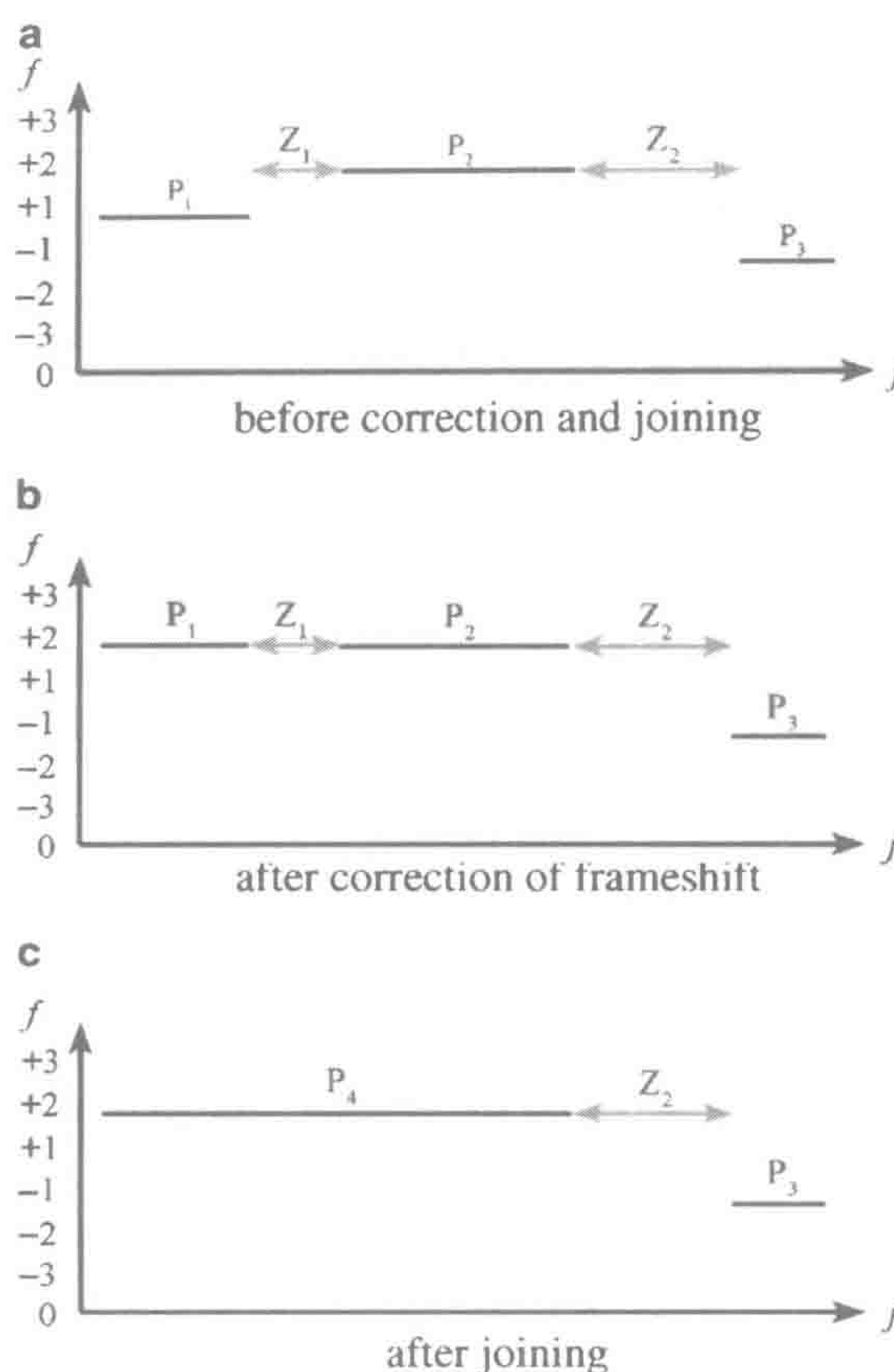


Fig. 4. (**a**) Three PCSs P_1, P_2, and P_3 on different reading frames. The distance between P_1 and P_2 as well as the distance between P_2 and P_3 is not more than 200 bp. Suppose there is a stop codon in frame + 2 in P_3, but not in P_1. Then the frame for P_1 is corrected, see subfigure (**b**). P_1 and P_2 in subfigure (**b**) are on the same reading frame. Suppose they are at distance at most 400 bp and z_1 does not contain a stop codon in frame + 2. Then P_1 and P_2 are joined to form P_4, see subfigure (**c**).

a synthetic metagenome allows us to quantify the success of *MetaGenomeThreader* in predicting coding sequences. The synthetic metagenome was constructed as follows: we collected complete genomes of two bacteria and of one archaea species. While the bacterial species are both dominant members of a maritime microbial community, the archaea species is an under-represented species (8). From each of the two bacteria we extracted a random region of 3,000 bp. From the archaea we extracted a random region of 2,000 bp. Due to the high density of genes, each of the extracted regions overlaps with or contains three or four genes. Six of the 11 genes are completely contained in the region and five are only partially covered. See Table 1, column 1–3 for details.

The three extracted regions were then separately fed into the ReadSim program (9), using its "454-mode." That is, ReadSim randomly extracts substrings of on average 120 bp from the

Table 1
The metagenome sequences used in the application

Species	Region	Coding sequence	Partial	Identified
Candidatus pelagibacter	4000...6999	- 3891... - 4832	×	√
NC_007205.1		- 4835... - 6247		√
		- 6249... - 6434		
		- 6444... - 7310	×	√
Vibrio cholerae	795000...797999	+ 794241... + 795380	×	√
NC_002505.1		+ 795485... + 795817		√
		+ 795839... + 797692		√
		+ 797707... + 798654	×	√
Pyrococcus horikoshii	172000...173999	- 171622...- 172662	×	√
NC_000961.1		- 172610...- 173815		√
		+ 173822...+ 174907	×	

We extracted substrings from two bacterial genomes (*C. pelagibacter* and *V. cholerae*) and one archaea genome (*P. horikoshii*). The second column shows the positions of the extracted regions of the genomes. The third column gives the positions and strand of the coding sequences relative to the start of the genome. Each region covers three or four coding sequences (according to the Genbank annotation of the genome) in the position range given in column 3. Column 4 marks the coding sequence, which are only partly covered. Column 5 marks the coding sequences, which were computed by *MetaGenomeThreader*.

extracted sequence regions and introduces a small number of insertions, deletions, and mismatches according to some error model specific for 454-reads. ReadSim produced 668 reads of total length 81,831 bp. This is a coverage of 10.23. These reads represent the synthetic metagenome.

The reads were then used as queries for a TblastX search delivering the local amino acid alignments. As a reference sequence set the NT-database (10) was used. It contains 21,789,657,596 bp. in 5,935,013 sequences. On a 2.1 GHz Duo Core Laptop TblastX used 27 h to complete, which gives an average running time of 145 s per read. The results of the TblastX run (i.e., the local alignments on the amino acid level) were stored on file (total file size: 62.9 mb) and processed by *MetaGenomeThreader*, using default parameters. This delivered 81 prediction of PCSs in about 15 min. The original 11 coding sequences were translated into proteins and then a TblastX run was performed to map the PCSs (as queries) to the proteins. It turned out, that 63 of the 81 PCSs (i.e., 78%) could be mapped to their original proteins (see Table 2 for a list) and 9 of 11 proteins were correctly predicted (see Table 1). In the successful cases, in spite of the short protein sequences derived from the PCSs, the target

Table 2
Mapping of the PCSs to the target proteins using TblastX

Target protein	Protein of the BLAST hit (Swissprot / TrEMBL-DB)	Counts of PCSs	Database subsequence length (in %)	Sequence identity (in %)	Protein sequence identity (in %)
tRNA isopentenyl-transferase	tRNA delta(2)-isopentenyl-pyrophosphate transferase	1	1×100	1×60 – 69	1×100
Probable periplasmic serine protease DO-like precursor	Probable periplasmic serine protease DO-like	2	2×80 – 89	2×100	2×100
Probable integral membrane proteinase	Probable integral membrane proteinase	2	1×70 – 79 1×80 – 89	1×90 – 99 1×100	2×100
Queuine tRNA-ribosyltransferase	Queuine tRNA-ribosyltransferase	3	1×40 – 49 2×50 – 59	1×70–79 2×90–99	1×63 1×81 1×100
Preprotein translocase subunit YajC	Putative uncharacterized protein	3	2×50–59 1×90–99	1×70–79 2×100	3×100
Protein export protein SecD	Protein-export membrane protein SecD (Putative uncharacterized protein)	29	1×30–39 2×40–49 1×50–59 4×60–69 7×70–79 4×80–89 7×90–99 1×100 1×> 100	(1×56) 1×50–59 4×70–79 6×80–89 5×90–99 13×100	(1×32) 1×89 27×100
Protein export protein SecF	Protein-export membrane protein SecF	1	1×97	1×100	1×100

(continued)

Table 2 (continued)

Target protein	Protein of the BLAST hit (Swissprot / TrEMBL-DB)	Counts of PCSs	Database subsequence length (in %)	Sequence identity (in %)	Protein sequence identity (in %)
DNA primase small subunit	DNA primase small subunit	12	1×40–49 2×50–59 1×80–89 1×90–99 7×100	1×70–79 1×80–89 3×90–99 7×100	1×100
DNA primase large subunit	DNA primase large subunit	10	1×40–49 1×50–59 2×90–99 5×100 1×> 100	2×70–79 8×100	1×100

Column 1 shows the name of the protein according to the annotation of the original genome section. Column 2 shows protein sequences obtained by a TblastXrun with the PCSs delivered by *MetaGenomeThreader*. The number of PCSs which could be assigned to the specified target proteins is shown in column 3. Column 4 shows the coverage of the BLAST hits on the protein sequences corresponding to a PCS. Column 5 shows the identity of the protein sequence alignment and in column 6 the identity of the target and the protein-sequence detected by Blastx is shown.

protein sequences could be identified with high probability, see Table 2 for details. The remaining 18 PCSs were incorrectly identified due problems with the detection of the reading frame, which is done by a greedy strategy that maximizes the length of the predicted gene. In cases, where the correct reading frame contains stop codons, other reading frames with longer predicted genes are favored by the greedy strategy. The algorithm could be considerably improved by a better strategy to detect the correct reading frame. The proteins that were not identified by *MetaGenomeThreader* are both the shortest proteins corresponding to DNA sequences of 186 bp and 172 bp, respectively. A complete description of the results can be found in (5).

4. Notes

When applying the gene finding method based on *MetaGenome Threader* one should note the following:

1. The input sequences (metagenome and reference sequences) are required to be in Fasta-format, a simple and widely used format for DNA and protein sequences.
2. TblastX needs to be run with reasonable parameters to balance sensitivity and specificity of the gene finding approach, always taking into account the available computing resources. If the TblastX parameters are too relaxed, it may run too long or may produce too large outputs which cannot be analyzed in reasonable time. If the TblastX parameters are too restrictive, it is unlikely that genes will be predicted.
3. The TblastX run is the bottleneck of the computation. When analyzing a complete metagenome, a computer cluster running TblastX is required to finish the analyses in reasonable time.
4. To speed up the computation, it is reasonable to restrict the application of TblastX to certain reference sequences and to exclude sequences, e.g., from a distant part of the phylogenetic tree. For example, in a bacterial metagenome project, one could exclude eukaryotic sequences.
5. As the TblastX output only contains protein alignments, *Meta GenomeThreader* needs access to the original DNA reference sequences. *MetaGenomeThreader* has two ways to obtain the access: either by online queries via http to the NCBI Genbank database or by scanning the original reference sequence in a batch mode. We recommend the second variant (option -k).
6. *MetaGenomeThreader* requires the protein alignments delivered by TblastX to be in XML-format. To achieve this, TblastX needs to be run with the format option -m 7.

References

1. Krause, L., Diaz, N.N., Bartels, D., Edwards, R.A., Pühler, A., Rohwer, F. et al. (2006) Finding novel genes in bacterial communities isolated from the environment. *Bioinformatics* 22(14), e281–e289.
2. Badger, J.H. and Olsen, G.J. (1999) CRITICA: coding region identification tool invoking comparative analysis. *Mol Biol Evol* 16(4), 512–524.
3. Moore, J.E. and Lake, J.A. (2003) Gene structure prediction in syntenic DNA segments. *Nucleic Acids Res* 31(24), 7271–7279.
4. Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A. et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437(7057), 376–380.
5. Metagenomethreader. http://www.zbh.uni-hamburg.de/mgth.
6. Genometools. http://genometools.org.
7. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. et al. (1997) Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Res* 25(17), 3389–3402.
8. Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.F., Rusch, D., Eisen, J.A. et al. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304(5667), 66–74.
9. Readsim – a simulator for sanger and 454 sequencing. http://www-ab.informatik.uni-tuebingen.de/software/readsim/welcome.html.
10. The blast databases. ftp://ftp.ncbi.nih.gov/blast/db/nt.gz downloaded on September 27, 2007.

INDEX

Wolfgang R. Streit and Rolf Daniel (eds.), *Metagenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 668, DOI 10.1007/978-1-60761-823-2, © Springer Science+Business Media, LLC 2010

P

Q

R

S

T

W

X

[General Information]
书名=元基因组学　　方法和步骤　　影印版
作者=（德）斯特赖特W．R．丹尼尔R．著
页数=341
SS号=14076076
DX号=
出版日期=2016．06
出版社=北京科学出版社